Structural and functional studies of proteins from the Hippo signalling pathway

Cherrett, Claire

Award date:
2011

Awarding institution:
University of Bath

Structural and functional studies of proteins from the Hippo signalling pathway

Claire Cherrett

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2011

ACKNOWLEDGEMENTS

First and foremost I would like to express my gratitude to my primary supervisor Stefan Bagby for his guidance, knowledge and expertise, and for taking the time to help ensure my work was published.

Thank you to both Stefan Bagby (again) and Jean van den Elsen for giving me the opportunity to carry out this work in such a friendly lab. Thanks also to Makoto Furutani-Seiki for the collaboration that resulted in the creation of my PhD project, and the BBSRC for my DTG funding.

My time here at the University of Bath would not have been as enjoyable, were it not for the excellent company of all lab members past and present. Particular thanks go to Abhishek Upadhyay for his many words of wisdom, support and instigating a host of thought provoking conversations.

For contributing their knowledge and expertise, I would like to thank Susan Crennell and Jean van den Elsen who provided a great deal of assistance during the X-ray crystallography saga. Our collaborator Rieko Ishima is gratefully acknowledged for providing expertise in NMR relaxation and for offering valuable advice and help with the interpretation of data relating to all things WW.

To my wonderful husband Scott: Thank you so very much for the love and support you have given me over these years through the good times and the bad.

The final words of appreciation go to my family and friends who have always been there for me. Thank you Mum; Dad; Joe; Emma; my Grandparents Margaret, Jim and Roy; and my Gran Stella – I wish you were still here to share this milestone with me.

ABSTRACT

The paralogous multi-functional adaptor proteins YAP and TAZ are nuclear effectors of the Hippo pathway, a central regulator of developmental organ size control, tissue homeostasis and tumour suppression. YAP/TAZ target the TEAD transcription factor family to promote cell survival and inhibit apoptosis. TEAD proteins contain a DNA-binding domain and a YAP/TAZ interaction domain. PCR analysis of medaka fish TEAD cDNA revealed the presence of alternative TEAD splice-forms with variations at the C-terminus of the DNA-binding domain. Structural analysis indicated the YAP-binding domain of TEAD proteins is folded and globular. NMR spectroscopy showed that the TEAD binding domain of YAP does not contain secondary structure. YAP and TAZ both contain WW domains, which are small protein-protein interaction modules. Two YAP isoforms are known, YAP1 and YAP2 that contain one and two WW domains, respectively. To date, only a single WW isoform of TAZ has been described. PCR analysis of medaka TAZ cDNA identified both single WW and tandem WW isoforms of TAZ. NMR spectroscopy was used to characterise structural, conformational, and peptide binding features of the tandem WW domains from YAP and TAZ. The YAP WW2 solution structure confirms that the domain has the canonical anti-parallel β-sheet WW fold. WW1 of YAP and both WW domains of TAZ undergo conformational exchange. The region linking the two WW domains is flexible and allows interaction of both WW domains with peptides containing single and dual PPxY binding motifs. In addition to YAP and TAZ, tandem WW domains are also present in the core and upstream Hippo pathway proteins Salvador and Kibra. Both proteins contain one atypical WW domain; the tandem WW domains of these two proteins are unstable. Understanding structure and function of Hippo pathway components could contribute to drug development and will also contribute to knowledge of protein folding and interactions.

ABBREVIATIONS

(l)	Long variant/construct
(s)	Short variant/construct
APS	Ammonium peroxodisulphate/persulphate
cDNA	Complementary DNA
CK1	Casein kinase 1
CKIP1	Casein kinase 2-interacting protein 1
Crb	Crumbs
DBD	DNA-binding domain
Dco	Discs overgrown
DNA	Deoxyribonucleic acid
Ds	Dachsous
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
dSTRIPAK	*Drosophila* Striatin-interacting phosphatase and kinase
DTT	Dithiothreitol
Dvl	Dishevelled
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial to mesenchymal transition
Ex	Expanded
FBP21	Formin-binding protein 21
FRMD6	FERM domain-containing protein 6
GST	Glutathione-S-transferase
HLH	Helix-loop-helix
Hpo	Hippo
HSQC	Heteronuclear single quantum coherence (spectrum)
IPTG	Isopropyl-1-thio-ß-galactopyranoside
LATS	Large tumour suppressor
LB	Luria-Bertani bacterial growth medium
Lft	Low-fat
Lgl	Lethal giant larvae
MAD	Multi-wavelength anomalous dispersion
Mats	Mob as tumour suppressor
Mer	Merlin
MIR	Multiple isomorphous replacement
MOB1	Mps1 binder 1
Mps 1	Monopolar spindle-1 kinase

MST	Mammalian Ste20-like kinase
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance (spectroscopy)
OPPF	Oxford Protein Production Facility
PATJ	PALS1-associated tight junction protein
PCR	Polymerase chain reaction
PDB	Protein data bank
PKC	Protein kinase C
PNK	Polynucleotide kinase
RASSF	Ras-association domain family
Rb	Retinoblastoma
RMSD	Root mean squared deviation
RNA	Ribonucleic acid
SAD	Single-wavelength anomalous dispersion
Sav	Salvador
SCRA	Sveinssons' chorioretinal atrophy
Sd	Scalloped
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Shh	Sonic hedgehog
SIR	Single isomorphous replacement
Su(dx)	Suppressor of deltex
TAD	Transactivation domain
TAZ	Transcriptional co-activator with PDZ binding motif
TBD	TEAD-binding domain
TEA	Transcription enhancer activation (domain)
TEAD	TEA domain containing transcription factor
TEF	Transcription enhancer factor
TEMED	N,N,N',N'-Tetramethylenediamine
TF	Transcription factor
Wts	Warts
WW	Protein domain containing two conserved tryptophan residues
YAP	Yes-associated protein
YBD	YAP-binding domain
Yki	Yorkie
ZO	Zonula occludens proteins

CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Introduction to this thesis

The body of work contained in this thesis comprises cloning, expression, purification and biophysical characterisation of protein domains that are involved in the Hippo signalling pathway. The Hippo pathway plays a vital role in development, regulating cell growth and organ size. Disruption of the pathway often leads to malignant tumourigenesis. Dividing key proteins from the pathway into putative domains allows for structural characterisation using biophysical techniques such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. Splitting proteins into smaller functional units also enables detailed analysis of protein-protein interactions.

In this chapter the Hippo pathway will be introduced and the proteins that were studied described in more detail. Chapter 2 outlines the general materials and methods used throughout the project. The interaction domains of the nuclear effector of the Hippo pathway, YAP, and its target transcription factor, TEAD are described in Chapter 3. The results presented in Chapter 4 were published in Biochemistry (Webb et al. 2011); the work characterises the WW domains from YAP and its homologue TAZ, and the WW domains' interactions with peptides containing the consensus binding-motif PPxY. Chapter 5 continues the thread of WW domains in the Hippo pathway, with NMR studies of tandem WW domains from the cytoplasmic proteins Salvador (Sav) and Kibra. The work is brought together in Chapter 6, a final discussion of the results and conclusions from the three preceding chapters.

1.2 An introduction to Hippo pathway proteins

1.2.1 Importance of the Hippo pathway

The Hippo pathway has emerged over the last decade as a key player in organ size regulation during development and tissue homeostasis throughout adult life. The signalling pathway also, perhaps unsurprisingly, plays a role in tumour suppression. Hippo pathway components were identified through loss of function genetic screens in *Drosophila melanogaster*; the *Hpo* gene, after which the pathway has been called, was named for the mutant overgrown head phenotype that resembled hippopotamus hide (Udan et al. 2003) (Figure 1.1).

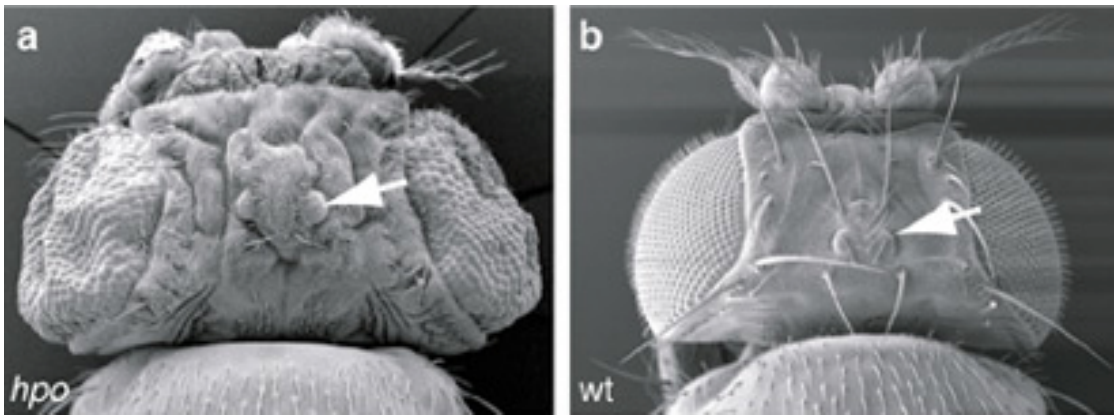


Figure 1.1. SEM images of (A) *Hpo* mutant with overgrown head phenotype and (B) wild type *Drosophila* heads. Arrows indicate the ocelli (eyes), which have developed normally. Image from Udan et al (2003).

The pathway has been shown to regulate organ growth in a coordinated manner such that when the nuclear effector YAP was upregulated in studies involving mice, the overall size of the liver increased significantly (Figure 1.2). Furthermore, after 8 weeks of YAP over-expression, tumourigenesis was observed and by 12 weeks the liver was riddled with hepatocellular carcinoma (HCC) (Dong et al. 2007). Makoto Furutani-Seiki (University of Bath) has shown that YAP is essential for the proper formation of organs in development of *Oryzias latipes* (medaka fish) embryos. The *hirame* mutant medaka embryos contain a null YAP mutation, with a distinctive phenotype characterised by a flat body, flattened neural tube and collapsed epithelial organs (Figure 1.3). Mutations in Hippo pathway components have also been observed in clinical cancer cases as well as *in vitro* cell lines. Therefore this signalling pathway is of interest both for potential regenerative medicine and cancer therapy uses, as well as elucidation of developmental control mechanisms.

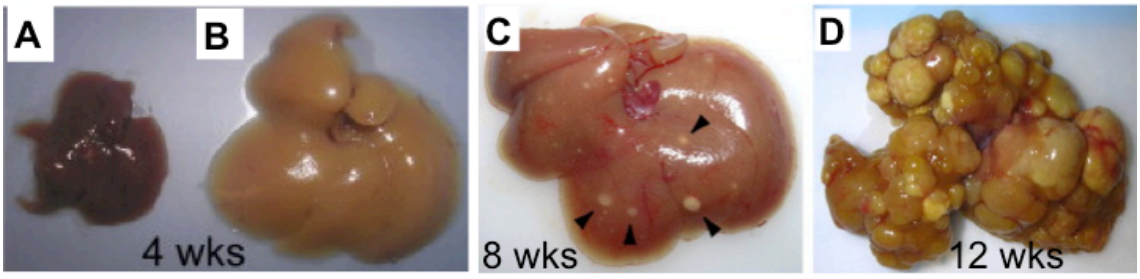


Figure 1.2. Transgenic mouse livers after (A) no YAP over-expression; (B) 4 weeks, (C) 8 weeks (with arrows indicating tumours) and (D) 12 weeks of YAP overexpression. Figure adapted from Dong et al (2007).

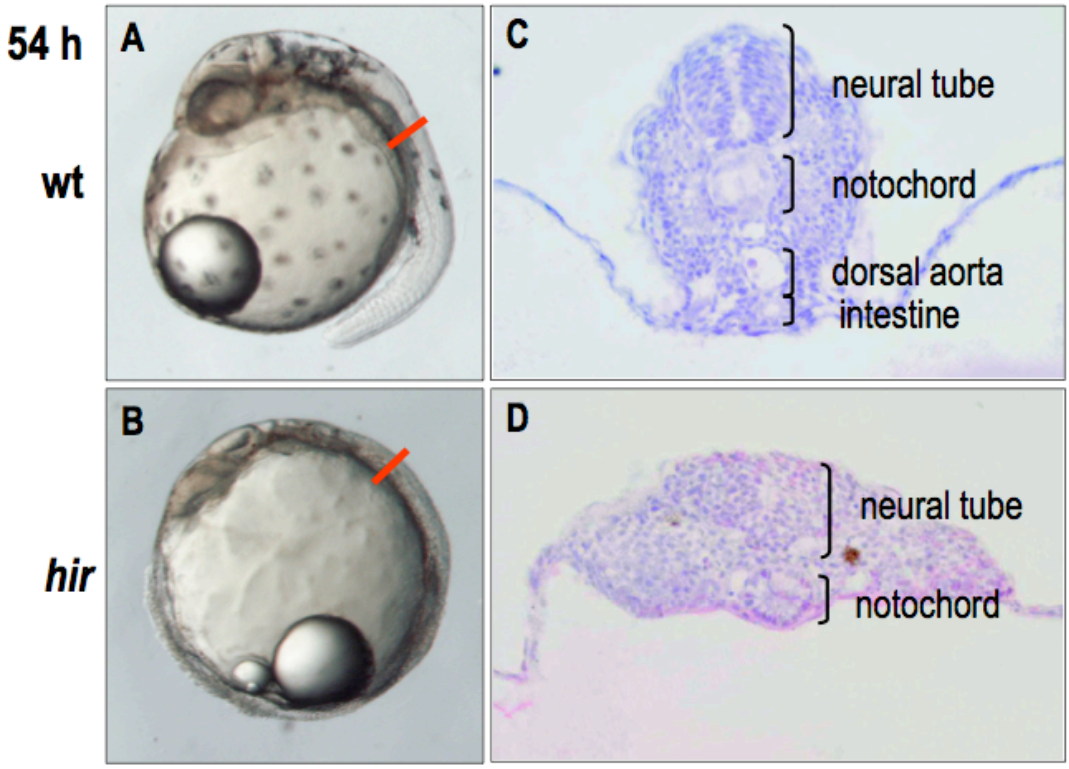


Figure 1.3 The YAP (*hir*) mutant vs wild-type medaka embryo phenotypes. (A, B) Lateral view of live embryos. (C, D) Frontal section of fixed embryos. (A, C) Wild type and (B, D) mutant embryos. The neural tube and somites are flat and the dorsal aorta and intestine are missing in the YAP mutant medaka embryos. Image provided by Makoto Furutani-Seiki (University of Bath).

1.2.2 Canonical Hippo pathway in Drosophila

The canonical Hippo pathway (Figure 1.4) is initiated in response to cell contact and extracellular matrix-induced tension (Dupont et al. 2011) through as yet unidentified receptors. In *Drosophila,* the most upstream component identified is the large atypical cadherin Fat, which is located at the sub-apical region of the plasma membrane during Hippo pathway activation and is associated with Dachsous (Ds) on neighbouring cells. The subsequent cellular signalling from Fat to the core kinase cassette requires phosphorylation of Fat by Discs overgrown (Dco) and interaction of Fat with Low-fat (Lft) in order to inhibit Dachs, an atypical myosin that inactivates the core kinase Warts (Wts). Fat also appears to activate the protein Expanded (Ex), but the exact process of this signal transduction is currently unclear (Badouel et al. 2009; Kango-Singh et al. 2009; Luo 2010; Sudol et al. 2010; Zhao et al. 2010; Zhao et al. 2010; Bao et al. 2011; Halder et al. 2011).

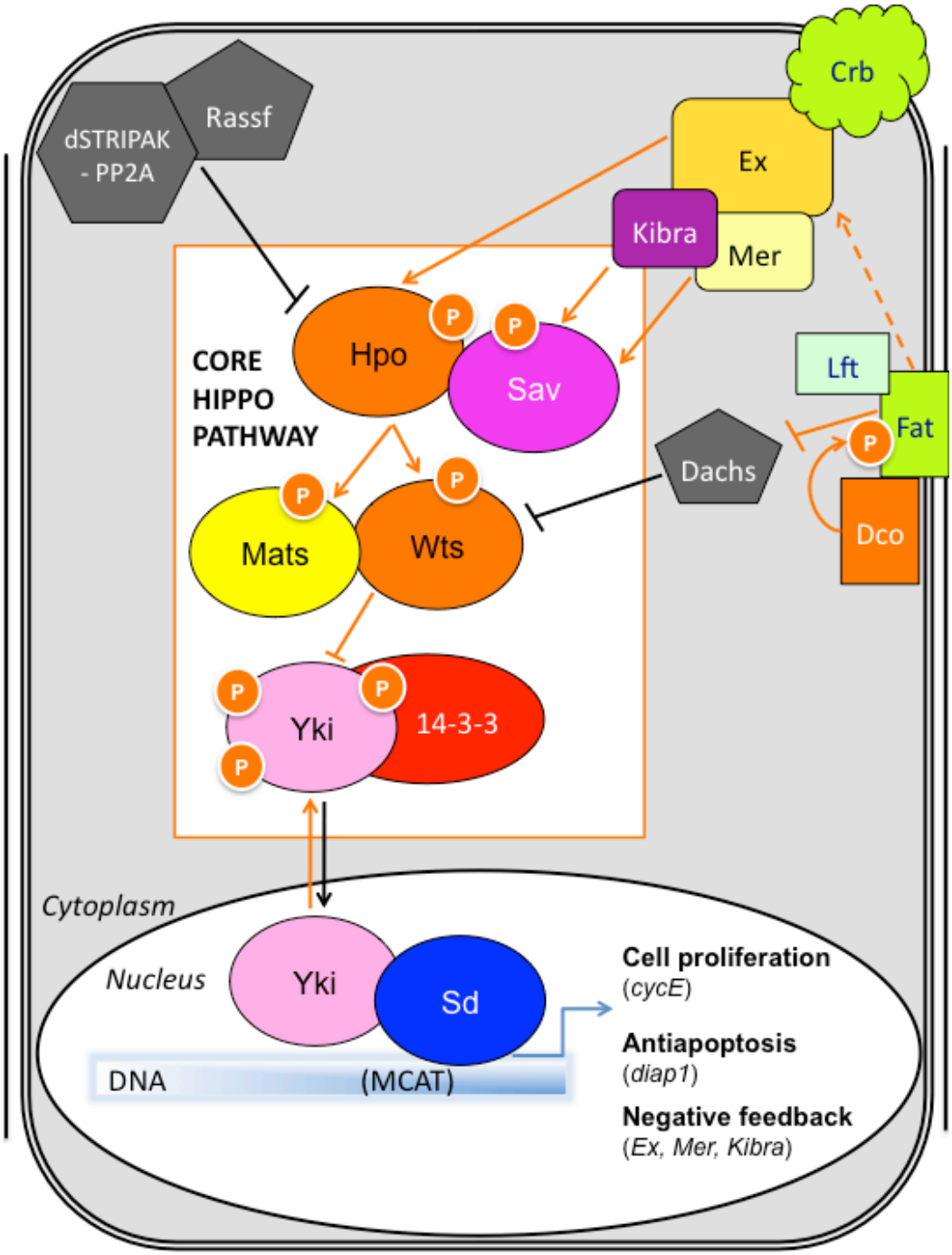


Figure 1.4 Representation of the Hippo pathway in *Drosophila*. Kinases and arrows that promote Hippo pathway activation are coloured orange; the core conserved Hippo kinase cascade is located within the orange box; proteins and arrows that inhibit the Hippo pathway are coloured in dark grey and black.

Another upstream regulator is the Crumbs (Crb) complex that is also located in the sub-apical region of the cell membrane (reviewed in (Parsons et al. 2010)). Crb recruits Ex to the plasma membrane, Ex then associates with Merlin (Mer), and the interaction is stabilised by binding to Kibra. This complex then binds to the core Hippo pathway

components Hpo and Sav through multiple interactions (Ex-Hpo, Mer-Sav, Kibra-Sav) and activates the core kinase cascade. Hpo is a kinase that autophosphorylates it's activation loop, then phosphorylates and forms an active complex with Sav. Hpo is then able to phosphorylate the kinase Wts and co-activator Mats (Mob as tumour suppressor). Wts subsequently phosphorylates the most downstream target of the Hippo pathway, Yorkie (Yki), at three sites, one of which enables 14-3-3 proteins to bind Yki and anchor it in the cytoplasm (Figure 1.4). When located in the nucleus Yki interacts with various transcription factors and acts as a transcriptional co-activator; therefore cytoplasmic sequestration renders Yki inactive (Badouel et al. 2009; Kango-Singh et al. 2009; Luo 2010; Sudol et al. 2010; Zhao et al. 2010; Zhao et al. 2010; Bao et al. 2011; Halder et al. 2011).

Inhibition of the Hippo pathway results in nuclear localisation of Yki, which enables Yki to co-activate the transcription factor Scalloped (Sd). Cell proliferation and survival genes are transcribed upon Yki-mediated activation of Sd. A negative feedback loop that leads to upregulated transcription of the Hippo pathway components Ex, Mer and Kibra is also initiated by Yki-Sd. One known mechanism of Hippo pathway inhibition is the formation of a complex between the *Drosophila* Ras-association domain family protein dRASSF (and Hpo, which competitively inhibits interactions between Hpo and Sav. dRASSF also associates with the phosphatase complex dSTRIPAK-PP2A that counteracts Hpo autophosphorylation (Ribeiro et al. 2010). dRASSF-mediated inhibition of the Hippo pathway appears to be under the control of atypical protein kinase C (aPKC) and Lethal giant larvae (Lgl) but the exact mechanism is not completely understood (reviewed in (Parsons et al. 2010)).

1.2.3 Hippo pathway in vertebrates

The downstream and core Hippo pathway proteins are conserved in vertebrates, though with more complexity (Figure 1.5). Many mammalian homologues of *Drosophila* Hippo pathway components have been shown to rescue mutant fly phenotypes providing evidence that the functionality of the pathway is conserved. The vertebrate orthologues of Sd are the Transcription enhancer activation domain (TEAD) transcription factors, TEAD1-TEAD4, which are also known as transcription enhancer factors (TEFs). The conserved TEA domain is a DNA binding region of around 70 residues, which is described in detail in section 1.2.4. The vertebrate orthologues of Yki are Yes-associated protein (YAP) and Transcriptional co-activator with PDZ binding

motif (TAZ). YAP was first discovered through its interaction with the SH3 domain of the c-Yes Src tyrosine kinase (Sudol 1994); however the role of YAP in Yes-signalling has not been extensively studied since the identification of this interaction. YAP and TAZ are highly similar in structure and function but can play different functional roles depending on cellular circumstances.

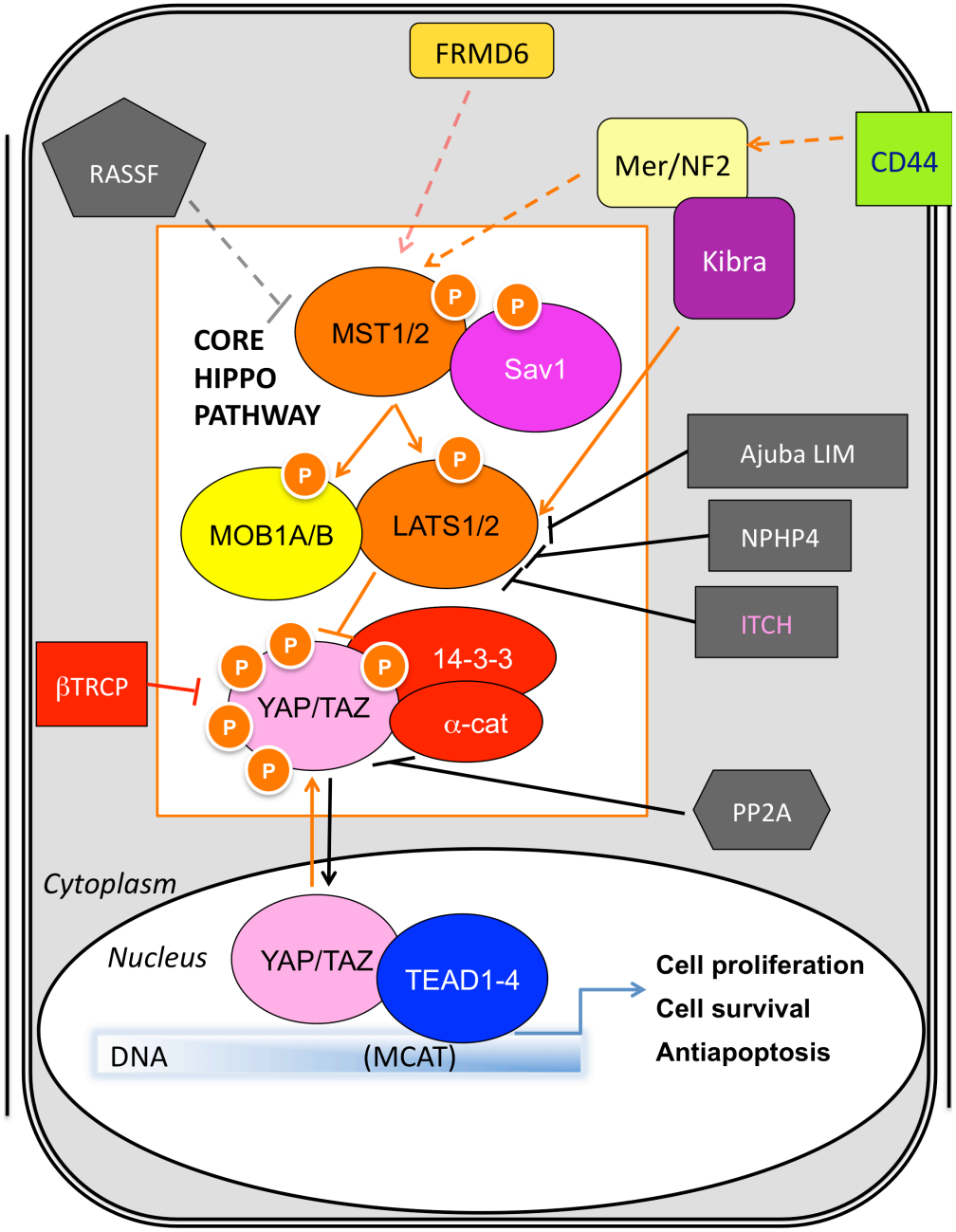


Figure 1.5 Representation of the Hippo pathway in vertebrates. Kinases and arrows that promote Hippo pathway activation are coloured orange; the core conserved Hippo kinase cascade is located within the orange box; proteins and arrows that inhibit the Hippo pathway are coloured in dark grey and black; proteins that contain WW domains are coloured or labelled in various shades of pink/purple.

The other vertebrate homologues of the Hippo pathway core kinase cassette are Mps1 (Monopolar spindle protein 1) binder-1 (MOB1) A and MOB1B (Mats in *Drosophila*), Large tumour suppressor (LATS) 1 and LATS2 (Wts in *Drosophila*), Sav1 (Sav in *Drosophila*), and Mammalian Ste20-like kinase (MST) 1 and MST2 (Hpo in *Drosophila*). Phosphorylation of YAP by LATS occurs at five sites bearing the HxRxxS motif resulting in the inhibitory interaction with 14-3-3. In mammals the YAP-14-3-3 complex is stabilised by α-catenin (Schlegelmilch et al. 2011). YAP phosphorylation by LATS can also lead to YAP degradation through CK1δ/ε mediated hyperphosphorylation and ubiquitination by the E3 ligase $SCF^{\beta TRCP}$ (Bao et al. 2011; Halder et al. 2011).

Most of the upstream *Drosophila* Hippo pathway proteins also appear to have counterparts in vertebrates (Figure 1.4) but the functional transmission of signals is less clear than in *Drosophila*. Mammalian Kibra cannot bind the apparent Ex homologue FERM domain-containing protein 6 (FRMD6), though there is some speculation as to whether this protein is a true homologue due to structural and functional differences (Genevet et al. 2010). Kibra is however able to interact with vertebrate Mer (a.k.a. NF2) (Genevet et al. 2010), and in addition to this it can bind directly to the LATS kinases, thereby regulating YAP phosphorylation (Xiao et al. 2011). The effect of mammalian Mer on MST1 and MST2 (Hpo homologues) has yielded conflicting data and so is also unclear. Crumbs (Crb), Fat4 (Fat) and Dachsous1 (Ds) localise in cell membranes in a similar manner to their *Drosophila* counterparts but have not yet been shown to affect FRMD6 (Ex), and a Dachs substitute has not been found. CD44, which can interact with Mer, has been suggested as another membrane bound regulator of the mammalian Hippo pathway (reviewed in (Bao et al. 2011)).

Mammalian PP2A antagonises the mammalian Hippo pathway by dephosphorylating YAP and acting in direct competition with the α-catenin-14-3-3 complex (Schlegelmilch et al. 2011). The relationship between MST (Hpo) and PP2A is more complex in mammals than in *Drosophila.* The mammalian MST seems to contribute to the stabilisation of PP2Ac rather than being inhibited by the phosphatase as occurs in *Drosophila* (Kilili et al. 2010). Mammalian RASSF proteins have been found to interact with MST, however the role of RASSF within the Hippo pathway is not clear (Avruch et al. 2011). Recently, other negative regulators of LATS have been elucidated, including the cilia-associated protein NPHP4 (Habbig et al. 2011) and the Ajuba LIM adaptor proteins (Das Thakur et al. 2010). The *Drosophila* counterpart of the Ajuba LIM proteins, djub, interacts genetically with the Hippo pathway. In addition to the

complexities regarding the function of upstream Hippo pathway components, several examples of crosstalk with other signalling pathways have been demonstrated (Table 1.1).

Table 1.1 Cross-talk between Hippo pathway proteins and other signalling pathways

Effect of Hippo signalling	Hippo pathway protein interaction
The Hippo pathway inhibits Wnt/β-catenin signalling (Varelas et al. 2010)	Direct interaction of TAZ with Dvl2 inhibits phosphorylation of Dvl2 by CK1δ/ε thereby preventing formation of the β-catenin destruction complex and resulting in β-catenin nuclear localisation and expression of Wnt target genes.
Hippo pathway inhibits BMP/TGFβ signalling (Alarcon et al. 2009; Varelas et al. 2010)	Direct interaction of YAP with phospho-Smad1 and YAP/TAZ with phospho-Smad2/3 retains Smads in the nucleus and leads to transcription of TGFβ/Smad target genes.
The Hippo pathway inhibits JAK/STAT signalling (Karpowicz et al. 2010)	Nuclear Yki induces transcription of cytokines that promote JAK/STAT signalling in response to injury.
The Hippo pathway inhibits Notch signalling (Reddy et al. 2010)	Active Yki inhibits the Notch ligand Delta which leads to Notch activation.
Sonic hedgehog (Shh) signalling inhibits the Hippo pathway (Fernandez et al. 2009)	Shh signalling up-regulates expression of YAP1 mRNA, and stabilises IRS1 (insulin receptor substrate 1) which acts as a nuclear retention factor for YAP. Shh signalling also results in decreased levels of phospho-LATS.
The Hippo pathway promotes FoxO signalling (Choi et al. 2009)	MST1 phosphorylates FoxO proteins leading to nuclear localisation and transcription of FoxO target genes.
PI3 kinase (PI3K)/Akt signalling inhibits the Hippo pathway (Yuan et al. 2010)	Akt phosphorylates MST1 (at T120), thereby preventing the kinase activity of MST1.
Retinoblastoma (Rb) (Nicolay et al. 2011; Tschop et al. 2011)	The transcription factor E2F is negatively regulated by Rb. E2F interacts with the Yki-Sd complex; therefore Rb inhibits E2F-Yki-Sd mediated transcription. In humans, LATS phosphorylates DYRK (dual-specificity tyrosine phosphorylation-regulated kinases), which leads to activation of the 'DREAM' complex and inhibition of E2F.
EGF-receptor signalling (Zhang et al. 2009)	YAP mediates transcription of the EGF-R ligand amphiregulin, leading to proliferation and migration of neighbouring cells.

1.2.4 TEAD transcription factors (the nuclear targets of YAP/TAZ)

The Hippo pathway monitors and controls the interaction of YAP and TEAD. The TEAD transcription factors are ubiquitously expressed, though each TEAD protein (TEAD1-TEAD4) occupies a slightly different niche with respect to tissue expression and developmental stage (Kaneko et al. 1997). TEADs induce transcription of genes that regulate cell proliferation, morphogenesis and differentiation, and so are crucial during development and tissue homeostasis (Chen et al. 2010). TEAD proteins have two discrete domains (Figure 1.6), the N-terminal DNA-binding domain (DBD) and the C-terminal YAP-binding domain (YBD). YAP is more complex, with several domains that mediate protein-protein interactions (Figure 1.6). The TEAD-binding domain (TBD) is located at the N-terminus of YAP. This domain of YAP seems to bind TEAD with high fidelity; to date no other protein interactions have been found to occur in the region.

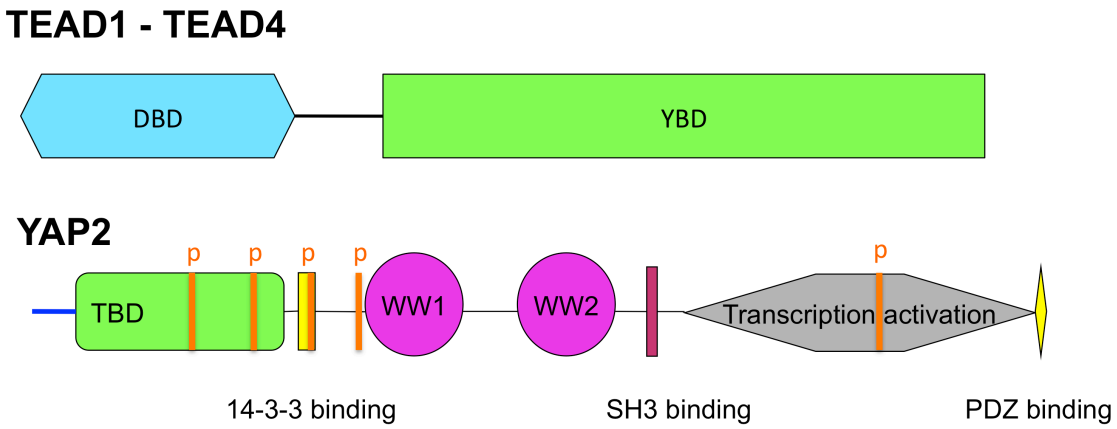


Figure 1.6 Representation of TEAD and YAP protein domains. The domain and linker lengths are to relative scales. The DNA binding region of TEAD TFs (blue) is approx 65-70 residues; the YAP-TEAD interaction domains of TEAD and YAP (green) are around 230 and 50 residues respectively; the WW domains of YAP are approx 40 residues and the inter-WW linker is around 20 residues. LATS phosphorylation sites (HxRxxS) (Hao et al. 2008) are highlighted in orange. The SH3 binding motif (maroon) is phosphorylated by Tyr kinases such as c-Yes.

The nuclear localisation signal (NLS) within the DBD of TEAD is vital to the activity of the TEAD-YAP or TEAD-TAZ complex. Sd is known to transport Yki into the nucleus as Yki/YAP lacks a NLS (Goulev et al. 2008). TEADs were also found to be responsible for the nuclear transport and retention of the YAP paralogue TAZ (Chan et al. 2009). The TEAD DBD can bind to a variety of M-CAT-like DNA sequences (the M-CAT motif is 5′-TCATTCCT-3′) in a fairly promiscuous manner (Anbanandam et al. 2006). M-CAT elements specifically bind TEF family proteins and are located in promoter-enhancer regions of various muscle-specific genes where they play important roles in muscle development (Yoshida 2008). There is no single mechanism of transcriptional activation by TEAD proteins; transcription is regulated by phosphorylation of TEAD, the presence of co-factors, interactions with other transcription factors and DNA binding proteins, accessibility of the M-CAT sequences to TEADs, and alternative splicing of TEFs (Yoshida 2008).

The solution structure of the human TEAD1 (TEF-1) A49S mutant DBD (which is also known as the TEA domain) (Figure 1.7) comprises a three-helix bundle in a homeodomain-like fold (Anbanandam et al. 2006). The first two helices are almost anti-parallel and the third helix lies across them both. The A49S mutation was made in the α1-helix (based on the *Drosophila* Sd DBD sequence) to improve solubility of the protein for use in NMR experiments. In a similar manner to homeodomain proteins, DNA binding is mediated by the third α-helix (H3) and the loop preceding it (L2); H1 and L1 do not bind directly but are necessary for full strength binding of TEAD to tandem M-CAT sites (Anbanandam et al. 2006).

YAP and TAZ are not the only nuclear co-factors of the TEAD transcription factors; gene specific transcription seems to be dependent on the interactions of TEAD with particular coactivators that presumably act to recruit the relevant transcriptional machinery. YAP and TEAD co-evolved, placing selection pressure on maintaining the interaction between the two proteins (Hilman et al. 2011). TEAD orthologues can be traced back to the ancient eukaryotic yeast *Saccharomyces cerevisiae*. YAP appeared later, the earliest example was found in the metazoan *Trichoplax adherens*; YAP was not identified in plants or yeasts (Hilman et al. 2011).

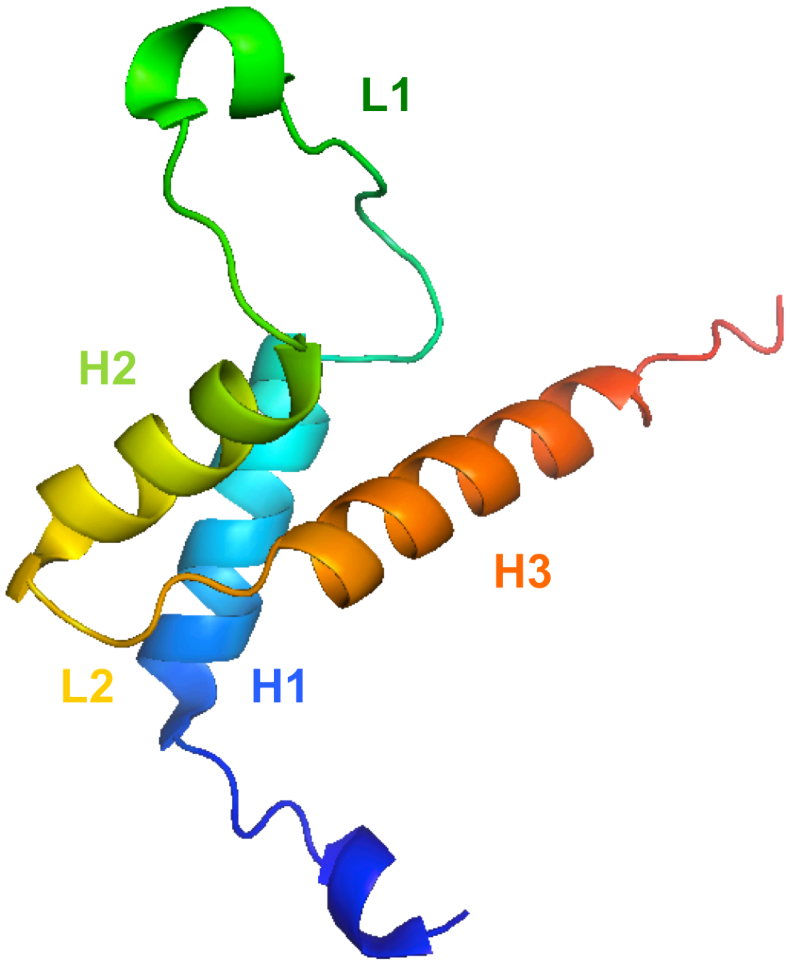


Figure 1.7 The TEAD1 A49S DNA-binding domain, with the N-terminus in blue and the C-terminus in orange (PDB ID: 2HZD (Anbanandam et al. 2006)). The image was made using PyMol (MacPyMOL 2009-2010).

The crystal structures of TEAD and YAP interaction domains from human TEAD2 (Tian et al. 2010), human TEAD1-YAP (Li et al. 2010), and mouse TEAD4-YAP (Chen et al. 2010) were published in 2010. The YAP-TEAD complex has been described as a hybrid transcription factor (Luo 2010) because individually the two proteins have no transcriptional activation capabilities. The Y421H mutation in TEAD1 that is present in human Sveinsson's chorioretinal atrophy (SCRA) was previously found to abrogate interactions with YAP and TAZ (Kitagawa 2007). As anticipated this residue is located in the TEAD-YAP interface where it forms a hydrogen bond with a Ser residue in YAP (S94 in human YAP, S79 in mouse YAP) (Fig 1.8e) (Chen et al. 2010; Li et al. 2010).

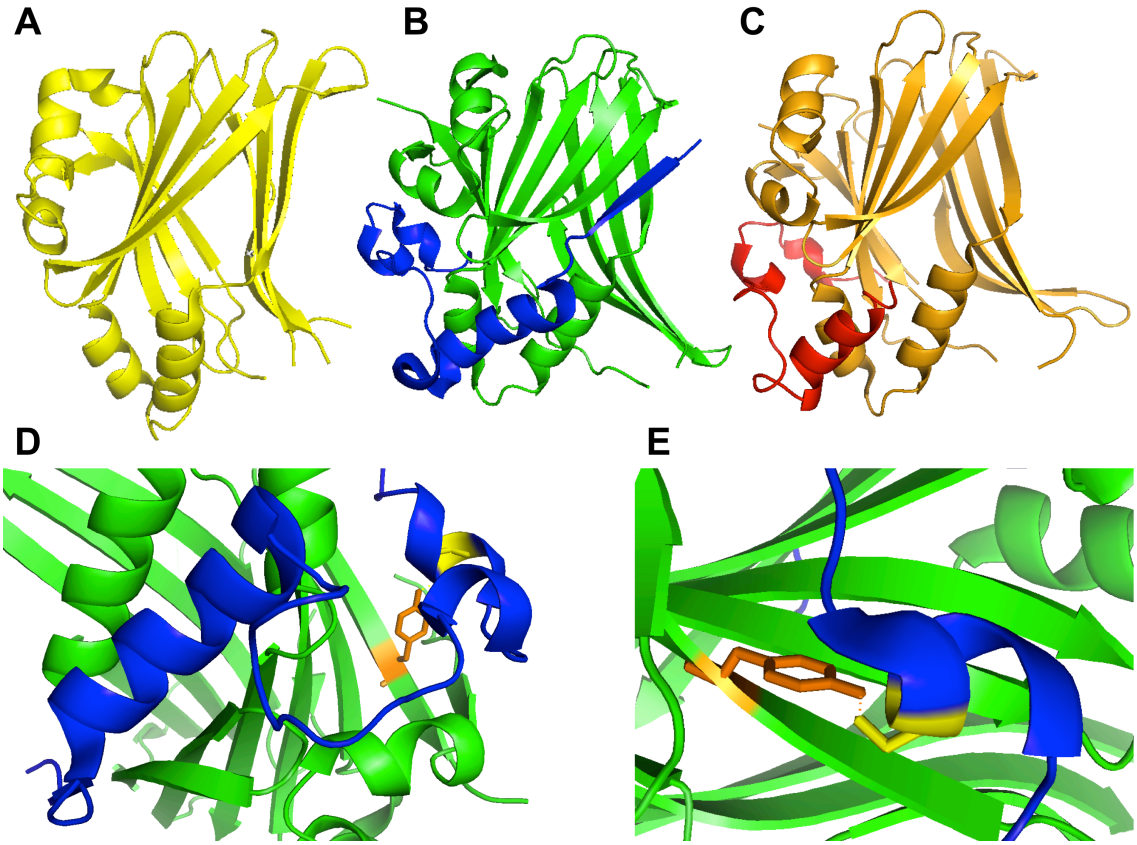


Figure 1.8 Crystal structures of (A) human TEAD2 (YBD) (PDB ID: 3L15), (B) human TEAD1 (YBD)-YAP (TBD) (PDB ID: 3KYS), and (C) mouse TEAD4 (YBD)-YAP (TBD) (PDB ID: 3JUA). YAP fragments are shown in (B, D, E) blue and (C) red. (D) The essential TEAD-binding helices of YAP comprise a longer helix that fits into a binding pocket within TEAD and two short helices C-terminal to a flexible loop that wraps around TEAD. (E) An essential H-bond forms between S94 of YAP (yellow) and Y421 of TEAD1 (orange). The images were made using PyMol (MacPyMOL 2009-2010).

The crystal structures of human TEAD2, human TEAD1-YAP and mouse TEAD4-YAP (Figure 1.8) show that the YAP-binding domains (YBD) of the TEAD proteins adopt an immunoglobulin-like β-sandwich conformation with the addition of two helix-turn-helix motifs. The TBD of YAP is natively unfolded but gains secondary structure upon interaction with TEAD (Fig. 1.8b,c). The TEAD structure is relatively unaffected by the interaction and YAP wraps around TEAD forming contacts in several places. The longer YAP peptide used in the human TEAD1-YAP structure (Fig. 1.8b) contains an N-terminal β-strand that is missing from the mouse TEAD 4-YAP structure; this strand does form contacts with TEAD but is not essential for the interaction. The 3 helices present in both structures (Fig 1.8) are all necessary for the interaction.

1.2.5 The many faces of YAP and TAZ

YAP and TAZ are multifaceted in both structure and function. The proteins are made up of modular domains that allow for interactions with many partners leading to a variety of, at times conflicting, functional roles. The modular structure of YAP is illustrated in Figure 1.6. TAZ is similar in structure to YAP, however TAZ lacks a proline rich region present at the N-terminus of human YAP and to date only a single WW domain isoform of TAZ has been described. The role of YAP/TAZ in cell proliferation was described in section 1.2.4, in addition to this function the proteins are also involved in epithelial to mesenchymal transition (EMT), differentiation, cell polarity, stem cell pluripotency, and apoptosis.

Although YAP was shown to require TEAD for nuclear localisation, YAP and TAZ are also known to interact with various other transcription factors including the pro-apoptotic p73 protein. In the case of p73, the PDZ binding motif at the C-terminus of YAP was shown to be essential for nuclear localisation and p73-YAP mediated transcription (Oka et al. 2009). Subsequent studies indicate this requirement is linked to interactions with the PDZ domain containing Zonula Occludens 1 and 2 (ZO) tight junction proteins (Oka et al. 2010; Remue et al. 2010). The PDZ-binding motif is also necessary for stabilising the YAP-p73 interaction (Oka et al. 2009). Although ZO proteins promote YAP-p73 activity, ZO2 was found to inhibit the transactivation activity of TAZ (Remue et al. 2010). This inhibition may be due to the formation of a ternary complex with an unidentified molecule.

In addition to the PDZ binding motif, the WW domains of YAP/TAZ (described in section 1.2.6) are also indispensable for the interaction with the consensus PPxY sequence of p73. The WW domains also mediate many interactions between YAP/TAZ and other binding proteins (Table 1.2) including the upstream Hippo pathway components LATS, Ex and Crb. The two isoforms of YAP (YAP1 and YAP2) contain one and two WW domains, respectively. The functional differences between isoforms are not completely clear but the presence of the second WW domain does appear to enhance the interaction of LATS1, which contains two PPxY motifs, with YAP (Zhang et al. 2008) and TAZ (Lei et al. 2008).

Table 1.2 Binding partners of YAP and TAZ

Binding protein	Description of binding protein	YAP/TAZ domain involved	Reference
HNRNPU	Heterogenous nuclear ribonuclear protein U binds to YAP and p73	**Proline-rich region at N-terminus of YAP**	(Howell et al. 2004)
TEAD	Transcription factor	**TBD**	(Vassilev et al. 2001)
14-3-3	Dimeric proteins that sequester YAP to the cytoplasm	**Phosphorylated 14-3-3-binding motif**	(Kanai et al. 2000; Basu et al. 2003)
Ex	Upstream Hippo pathway FERM-domain protein	**WW domain**	(Badouel et al. 2009)
LATS1/2	Hippo pathway Ser-kinase	**WW domains**	(Lei et al. 2008; Zhang et al. 2008)
p73	Pro-apoptotic transcription factor	**WW domain and PDZ-binding motif**	(Strano et al. 2001; Oka et al. 2009)
WBP1/2	WW domain-binding proteins	**WW domain**	(Chen et al. 1995)
ASPP1/2 (p53BP)	Apoptosis-stimulating protein of p53 protein family	**WW domain**	(Espanel et al. 2001; Liu et al. 2011)
Runx1/2	Transcription factors with Runt DNA-binding domain	**WW domain**	(Yagi et al. 1999; Zaidi et al. 2004; Hong et al. 2005)
Smad1/7	Transcription factors regulated by the TGFβ BMP signalling pathway	**WW domains**	(Ferrigno et al. 2002; Alarcon et al. 2009)
ErbB4	Receptor Tyr-kinase that contains a cleavable cytoplasmic fragment	**WW domains**	(Komuro et al. 2003)
Dvl2	Dishevelled polarity protein involved in Wnt signalling	**WW domain**	(Varelas et al. 2010)
AMOT	Angiomotin part of the Crumbs complex	**WW domains**	(Chan et al. 2011)
PRGP2	Proline rich membrane Gla protein	**WW domain**	(Kulman et al. 2007)
PEBP2	Polyoma enhancer binding protein 2 transcription factor	**WW domain**	(Yagi et al. 1999)
NFE2 (p45)	Haemopoietic transcription factor-2	**WW domains**	(Gavva et al. 1997)
c-Yes	Tyr-kinase	**SH3-binding motif**	(Sudol et al. 1995)
NHERF (EPB50)	Recruits YAP and TAZ to plasma membrane	**PDZ-binding motif**	(Mohler et al. 1999; Kanai et al. 2000)
ZO1/2	Zonula occludens tight junction proteins	**PDZ-binding motif**	(Oka et al. 2010; Remue et al. 2010)

It is tempting to conclude that the nuclear activity of YAP is wholly dictated within the cytoplasm. Just as cytoplasmic proteins negatively regulate YAP/TAZ by phosphorylation and sequestration, evidence suggests that YAP and TAZ are transported from the cytoplasm into the nucleus by the transcription factors and co-factors that complete the active complexes. However, given the sheer number of identified (and potentially unidentified) binding partners of YAP/TAZ (some of which are in Table 1.2) the process is likely to be more complex. Therefore it is important to understand the structure of YAP/TAZ domains and how the domains behave during protein-protein interactions, both individually and together.

1.2.6 WW domains within the Hippo pathway and beyond

WW domains are the smallest known protein-protein interaction modules at around 35-40 residues long and are so named because of two conserved tryptophans (Trp) 20-22 residues apart. The domains also contain a conserved proline (Pro) two residues C-terminal to the second Trp. WW domains typically fold into a twisted three-stranded β-sheet. The first Trp is essential for domain folding and is located on the first β-strand within the hydrophobic core (Koepf et al. 1999). The second Trp contributes to ligand interactions and is located in the aromatic binding pocket. WW domains are categorised into five groups (I-V) according to their cognate ligand (Ball et al. 2005); all WW domains bind Pro-rich motifs (Table 1.3).

Table 1.3 WW domain groups and binding motifs. Adapted from Ball et al (2005)

Classification	Consensus binding motif	Example WW domain protein
Group I	PPxY	YAP
Group II	PPLP	FBP-11
Group IIIa	(p/ϕ)P(p/g)PPpR	FE65
Group IIIb	(p/ϕ)PP(R/K)gpPp	FBP-21
Group IV	(S^p/T^p)P	PIN1
Group V	(p/ϕ)PPPPP	PRP40

x any residue; ^{p} residue is phosphorylated; lower case letters signify favoured but not conserved residues; ϕ hydrophobic residues.

The Hippo pathway is rich in WW domains. YAP contains 1 or 2 WW domains that are used during interactions with many binding partners (Table 1.2) that contain 1 or 2

PPxY motifs. The upstream Hippo pathway proteins Sav and Kibra contain two WW domains each. ITCH E3 ubiquitin ligase, that was recently found to induce proteasomal degradation of LATS, also contains multiple WW domains (Salah et al. 2011). The functionality of WW domains extends far beyond the Hippo pathway. To date, almost 100 WW domains and around 1900 PPxY motifs have been identified within the human proteome (Sudol et al. 2010). Several diseases have been attributed to aberrant WW domain function including Liddle's syndrome (Nedd4-EnAC interaction) (Staub et al. 1996), Alzheimer's disease (Pin1-Tau, FE65-APP) (Sudol et al. 2001), and Duchenne's Muscular Dystrophy (dystrophin-β-dystroglycan-caveolin) (Ilsley et al. 2002).

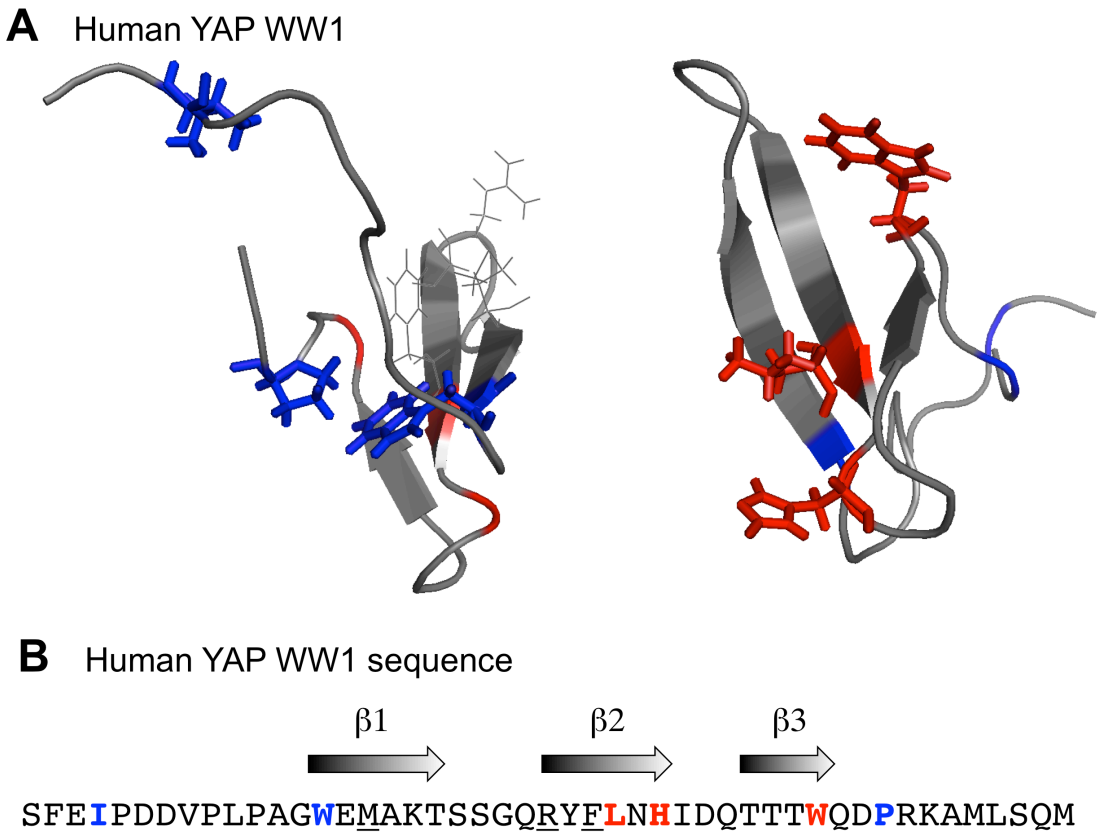


Figure 1.9 (A) Average solution structure of a YAP WW1 bound to a peptide (not shown) (PDB ID: 1K9R (Pires et al. 2001)). (B) Sequence of human YAP WW1 with locations of β-strands illustrated with arrows. Residues shown in blue are involved in folding and stability; other residues that contribute to the hydrophobic core are shown in (A) grey or (B) underlined. Residues involved in peptide binding and binding specificity are highlighted in red.

The WW domain was first identified and characterised in YAP1 (previously known as YAP65) (Sudol 1994). Subsequently the solution structure of YAP WW1 in complex with the peptide GTPPPPYTVG was published (Macias et al. 1996); this structural characterisation also showed that an Ile N-terminal to the core WW domain was necessary to stabilise the fold (Figure 1.9). Since the structure of YAP WW1 was revealed as the first example of a WW domain fold, it has been the subject of several other structural studies.

The previously mentioned property of stability being attributed to the first W and binding to the second W in WW domains was studied using YAP WW1 (Koepf et al. 1999). Espanel and Sudol (1999) showed that mutating the Leu and His in the second β-strand of the group I YAP WW1 domain to Trp and Ile switched the ligand binding preference of the domain to that of the group II WW domain-containing protein FE65. These residues (L and H in the β2 strand) are known to interact with the Y of the PPxY motif favoured by group I WW domains. Pires et al (2001) mutated residues of YAP WW1 that lead to formation of a more stable WW-PPxY complex. As observed in the group I to group II WW domain study, it was mutation of the β2 Leu (to Lys in this instance) that resulted in altered peptide-binding affinity. In the same study (Pires et al. 2001) a minimal peptide ligand consisting of acetyl-PLPPY was identified. A molecular dynamics simulation study supported previous observations that YAP WW1 requires the unstructured regions N- and C- terminal to the core β-sheet to maintain the WW domain fold and that ligand binding further stabilises the structure (Ibragimova et al. 1999). Figure 1.9 illustrates the WW domain fold of YAP WW1 and highlights the key residues involved in folding and binding.

YAP WW1 was the first of many WW domains to be studied using NMR spectroscopy; each published WW domain solution structure contains the common three-stranded β-sheet fold, however not all WW domains are natively folded. For example the WW3-WW4 tandem WW repeat from the E3 ubiquitin ligase Nedd4 family protein Suppressor of deltex (Su(dx)) is partially unfolded, and requires the interaction between WW3 and PPxY peptide to stabilise the structure of both WW3 and WW4 (Fedoroff et al. 2004). Another unusual feature of Su(dx) WW4 is a substitution of the second Trp by Phe (so it could be called a WF domain). The ‘WF’ domain of Su(dx) cannot bind the PPPY motif that was used in studies of YAP WW1 but is capable of binding the PPSY motif present in the Notch protein, though with a lower affinity than its typical WW counterparts (Jennings et al. 2007). The E3 ubiquitin ligase Smurf2 also contains a ‘WF’ domain (WW3). In the case of Smurf2, the ‘WF’ domain

can bind the PPPY motif of Smad7 but it also requires the presence of six residues C-terminal to the motif that stabilise the complex through additional interactions with the WW domain (Chong et al. 2006). As was observed for Su(dx), the Trp to Phe substitution in Smurf2 reduces the binding affinity for cognate ligand in comparison to a typical WW domain.

Tandem WW domains are common in Hippo pathway proteins (YAP, TAZ, Sav, Kibra, ITCH). WW domain doublets from a range of other proteins have been characterised using NMR spectroscopy. Solution structures of peptide-bound Su(dx) WW3-WW4 (PDB ID: 1TK7) (Fedoroff et al. 2004) and unliganded FBP21 (PDB ID:2JXW) (Huang et al. 2009) reveal a lack of secondary structure in the inter-domain linker. Conversely the linker between WW domains in the yeast splicing factor Prp40 forms a rigid α-helix that fixes the orientation of the WW domains (Wiesner et al. 2002). The inter-domain linker is a key factor in the regulation of interaction specificity for the Smurf ubiquitin ligases. The linker length varies between different isoforms of Smurf1/2; the isoforms with shorter linkers have higher affinity with peptide ligand due to cooperation between WW domains (Chong et al. 2010). The solution structure of Smurf1 WW1-WW2 with phospho-Smad1 peptide reveals a linker consisting of helix-turn-loop that allows for formation of a more stable ternary complex with CKIP1 (casein-kinase 2-interacting protein 1) (Aragon et al. 2011).

WW domains are key interaction modules within the Hippo pathway and beyond. Although many WW domains have been studied and despite their small size, the exact factors that determine folding and interactions are still not completely predictable. Tandem WW domains are common in the Hippo pathway and other important cell regulatory proteins, and each tandem WW domain to date exhibits different structural and functional characteristics. Therefore it is important to understand how the WW domain modules fold and function in proteins of the Hippo pathway for potential therapeutic modifications and drug targeting in the future.

1.2.7 Salvador, the WW domain-containing core Hippo pathway protein

Salvador (Sav/Sav1) was first identified in human and mouse in 2000 through an EST search for novel WW domain containing proteins; at that time the protein was named WW45 (Valverde 2000). As mentioned previously (Section 1.2.2 and 1.2.3), Sav is integral to the canonical Hippo pathway. It is a scaffold protein that contributes to multi-protein complex formation through its modular interaction domains. The N-terminus of Sav is not predicted to contain structural elements, two WW domains are

present in the middle of the protein, and a coiled-coil region is located at the C-terminus (Figure 1.10). The coiled-coil region was later shown to be a novel interaction domain common to, and allowing complex formation between, Sav, Rassf and Hpo; it was thus called a SARAH domain (Scheel et al. 2003). The solution structure of the SARAH domain from human MST1 (Hpo) is in the PDB (2JO8), but the Sav SARAH domain remains undetermined. In addition to the defined domains within Sav, a nuclear localisation signal (NLS) sequence was identified N-terminal to the WW domains, and two endoplasmic reticulum membrane retention signals are present at either end of the protein.

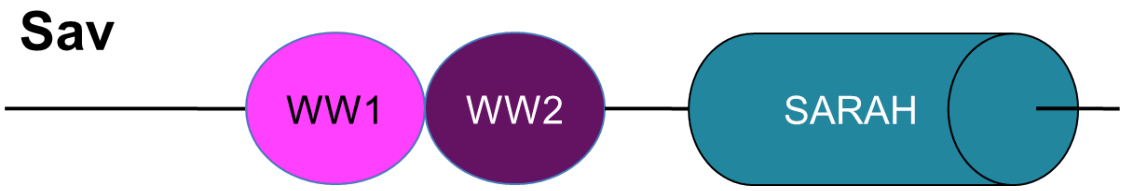


Figure 1.10 Representation of structured Sav domains. WW1 (pink) is a typical group 1 WW domain. WW2 (purple) is an atypical 'WY' type WW domain. The C-terminal SARAH domain is a coiled coil that can homo- or hetero-dimerise with the SARAH domains of RASSF and MST proteins. Inter-domain random coil regions are to relative scales in comparison to the structured domains; there is no inter-WW linker.

The first WW domain of Sav belongs to the group I classification; the solution structure of mouse Sav WW1 is in the PDB (2YSB). The second WW domain of Sav is unusual in that Y replaces the second conserved W (a 'WY' domain). Mouse Sav WW2 is also atypical because it exists as a homodimer in a β-clam like formation with the interface obscuring what would normally be considered as the ligand binding site (Ohnishi et al. 2007) (Figure 1.11b). Homodimerisation of WW2 may contribute to the role of Sav as a scaffold protein, allowing for the formation of larger protein complexes through interactions between Sav WW1-LATS and Sav SARAH-Rassf/Hpo SARAH.

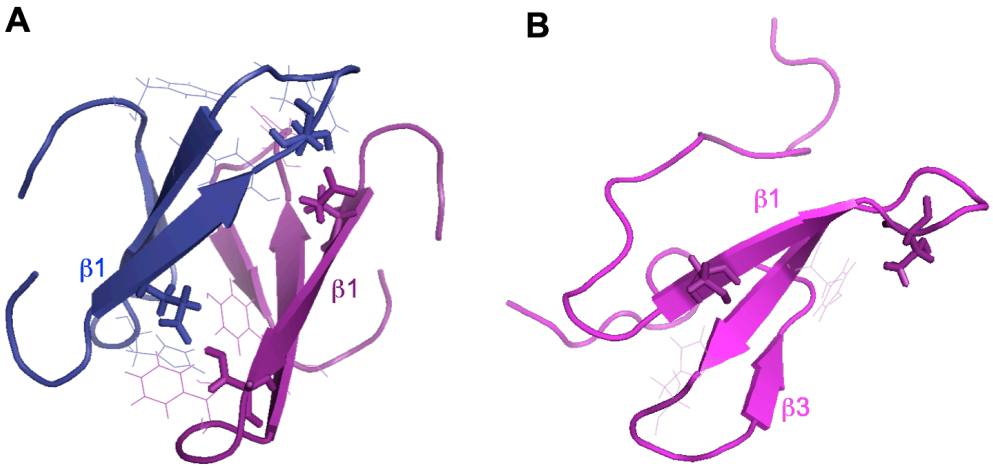


Figure 1.11 Solution structures of mouse Sav (A) dimeric WW2 (blue and purple to distinguish the monomers) (PDB ID: 2DWV) and (B) monomeric WW1 (magenta) (PDB ID: 2YSB). (A) The polar residues Glu and Ser that are located at the N- and C- termini, respectively, of the WW2 β1 strand contribute to the dimerisation and are shown in thick lines. The four aromatic residues (F,Y,H,Y) that also play a role in dimerisation are shown in thin lines. The equivalent residues in (B) WW1 to the Glu and Ser of WW2 (Ser and Thr) are shown in thick lines and only two equivalent aromatic residues are present in WW1 (thin lines). The images were made in PyMol (MacPyMOL 2009-2010).

1.2.8 Kibra, the upstream WW domain Hippo pathway protein

Kibra was first described in 2003 (Kremerskothen et al.), after being identified in a yeast two hybrid human brain cDNA library screen with a fragment of human Dendrin as bait. The protein was named for its predominant expression in kidney and brain. Kibra has a molecular weight of 125 kDa and comprises two WW domains at the N-terminus, followed by a predicted unstructured region that contains a nuclear localisation signal, a C2 domain, Glu-rich region, PKCζ-binding domain and a C-terminal PDZ-binding motif (Figure 1.12).

Kibra has several known interaction partners outside of the Hippo pathway including PKCζ/PKMζ, Dendrin and Synaptopodin, all of which are involved in dendritic cytoskeleton organisation. Through these interactions Kibra appears to play a key role in cognitive function and memory. Kibra has also been linked to Alzheimer's disease. Another known binding partner is PATJ (PALS1-associated tight junction protein). This protein is part of the Crumbs complex that is linked to the Hippo pathway in *Drosophila*, but is better known for regulating cell polarity and migration (reviewed in (Schneider et al. 2010)).

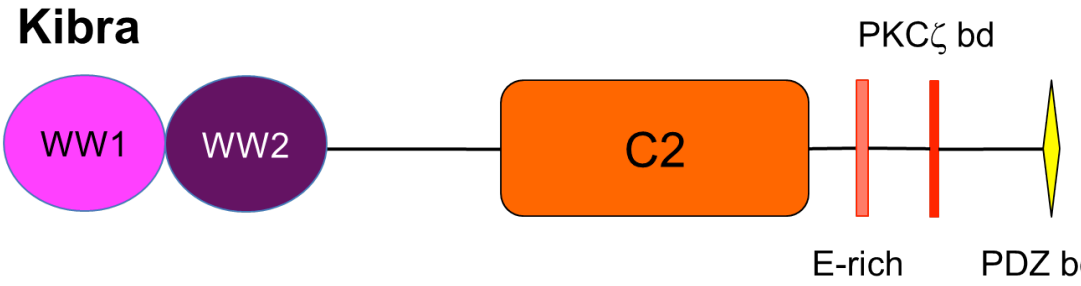


Figure 1.12 Representation of structured domains from Kibra. WW1 (pink) is a typical group 1 WW domain and WW2 (purple) is an atypical 'WY' type WW domain; there is no inter-domain linker. The role of the C2 domain (orange) in the Hippo pathway is currently unknown. The C-terminal region of Kibra is unstructured but contains a Glu-rich region (salmon pink), a PKCζ binding-motif (red) and a PDZ binding-motif (yellow).

The C-terminal interaction regions consist of short sequence motifs without any predicted secondary structure. A crystal structure of the Kibra C2 domain is in the PDB (2Z0U) but has not been published as part of a peer reviewed article. C2 domains are involved in Ca^{2+}-dependent phospholipid binding and are frequently associated with signalling molecules. The Kibra C2 domain is an 8 stranded β-sandwich (Figure 1.13) with a structure typical of these domains, one example of which is the first C2 domain of synaptotagmin (Sutton et al. 1995). Unlike synaptotagmin, however, the crystal structure of Kibra C2 does not contain a Ca^{2+} or any other metal ion.

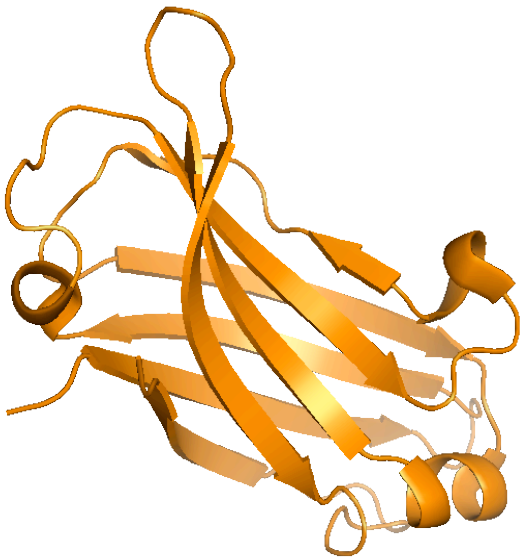


Figure 1.13 Crystal structure of the Kibra C2 domain (PDB ID: 2Z0U), the orange colour corresponds to the C2 domain representation in Figure 1.12. The image was made using PyMol (MacPyMOL 2009-2010).

No structural studies have been carried out on the WW domains of Kibra to date. The first WW domain of Kibra is predicted to be a typical group I WW domain. The second WW domain is unusual in that the second Trp is replaced with an Ile, and as might be expected given this substitution, WW2 is incapable of binding to peptides with the PPxY motif (Kremerskothen et al. 2003). The WW domains are of particular importance for the role of Kibra in the Hippo pathway. Upon deletion of the two WW domains the interaction of Kibra with LATS is abrogated, leaving LATS susceptible to ubiquitination and subsequent proteasomal degradation (Xiao et al. 2011).

1.3 Aims

The work described in this thesis is intended to resolve the structural features and protein-protein interaction properties of domains from five proteins within the Hippo pathway at the atomic level using NMR spectroscopy and X-ray crystallography. The key aims are to successfully clone target gene regions into suitable expression vectors for bacterial expression, purify sufficient quantities of homogeneous protein, prepare high quality protein samples, and produce structural and interaction data using NMR spectroscopy and X-ray crystallography.

The proteins and domains that feature are the TEAD transcription factors (DBD, YBD), YAP (TBD, WW domains) and TAZ (WW domains), Sav (WW domains) and Kibra (WW domains). Through studying the structure and binding properties (function) of these protein domains, better understanding of protein folding in general will be gained. Also, structural properties of Hippo pathway components will be revealed that may contribute to therapeutic targeting and drug design in the future.

CHAPTER 2: GENERAL MATERIALS AND METHODS

This chapter describes the materials and methods that were used throughout the project and are relevant to all subsequent results sections. Detailed experimental conditions for each construct will be described in the 'Materials and Methods' section of Chapters 3, 4 and 5 accordingly.

2.1 Chemicals and reagents

Where specific standard chemicals are not included in Table 2.1 it can be assumed that they were purchased from the company (listed next to 'chemicals') with the most competitive pricing.

Table 2.1 Sources of reagents and chemicals

REAGENT/ CHEMICAL	COMPANY
^{13}C-labelled D-glucose 99 atom %	Isotec
^{15}N-labelled NH_4Cl	Isotec
2-mercaptoethanol	Fluka
2-propanol	Sigma Aldrich
40% 29:1 acrylamide/bis-acrylamide	Fisher Scientific
Agarose	Melford
Antibiotics (kanamycin, ampicillin, chloramphenicol)	Melford
Ammonium peroxodisulphate	BDH/Sigma Aldrich
Brilliant blue R250	Fisher Scientific
Bromophenol blue	BDH
Chemicals (not otherwise mentioned)	Melford/Sigma Aldrich/Fisher Scientific
'Complete mini protease inhibitor cocktail' tablets	Roche
D-glucose	BDH
Deuterium oxide 99.9 atom %	Goss Scientific Instruments
Deoxynucleotide triphosphate set, PCR grade	Roche
DL Dithiothreitol	Sigma Aldrich
DNAse I	Ambion

Ethylenediamine tetraacetic acid (EDTA)	Fisher Scientific
Ethidium bromide (10 mg/ml solution)	Fisher Scientific
Glycerol	Fisher Scientific/Sigma Aldrich
Glycine	Sigma Aldich
HRV 3C protease	Novagen
Imidazole	Acros Organics
Isopropyl-1-thio-ß-galactopyranoside (IPTG)	Melford
KOD Hot Start polymerase	Novagen
Luria Bertani Agar, Miller	Fisher Scientific
MEM Vitamin (100x) solution)	Sigma Aldrich
Miniprep kits	Stratagene/Fermentas/Promega
PBS tablets	Oxoid
PCR/Gel clean-up kits	Stratagene/Fermentas/Promega
PEG 4000	Fluka
Phusion High-fidelity DNA polymerase	NEB
Primers	MWG/Invitrogen
Reduced glutathione	Acros organics/BDH
Restriction endonucleases (HF and Fast-digest)	NEB
Sodium citrate	Melford
Sodium dodecyl sulphate	BDH
Solvents (EtOH, MeOH, acetone)	Fisher Scientific/Sigma Aldrich
T4 DNA ligase	NEB
T4 PNK β	NEB
TaKaRa PCR kit	AMV
Taq DNA Polymerase	NEB
N,N,N',N'-Tetramethylenediamine (TEMED)	Sigma Aldrich
tris(hydroxymethyl)aminomethane (tris base)	Melford
Triton-X100	Fisher Scientific/Sigma Aldrich
TRIzol®	Invitrogen
Trypsin from bovine pancreas	Sigma Aldrich
Tryptone	Melford
Yeast extract	Melford

2.2 Plasmids, cell strains and media

2.2.1 cDNA plasmids for use as PCR templates

pBlueScript SK-. These vectors contain an ampicillin resistance gene and inserted sequences are under the regulation of the T7 promoter. Makoto Furutani-Seiki (University of Bath) was kind enough to provide EcoRV digested pBluescript SK- and a plasmid containing medaka K-Cab strain YAP cDNA.

Sav cDNA. Nic Tapon (Cancer Research UK) kindly supplied *Drosophila* and human Salvador cDNA plasmids.

Kibra cDNA. Human Kibra cDNA plasmids were a gift from Joachim Kremerskothen (Universität Münster).

2.2.2 Expression plasmids

Vanderbilt vectors (pSV…/pBG…) were supplied by Vanderbilt University Structural Biology Center; these plasmids are derived from pET vectors and are only for use in bacterial expression systems. The Oxford Protein Production Facility (OPPF), UK (Berrow, Alderton et al. 2007) supplied the pOPIN vectors; these vectors are designed for use in *E. coli*, insect and mammalian expression systems and are based on the pTriEX2^{TM} vectors (Novagen). Initially Vanderbilt vectors were used as they have been tried and tested within the group and are known to produce high protein yields under optimal conditions. The TEAD protein constructs (and YAP TBD) were largely insoluble when produced by the Vanderbilt vectors. Various protein domains were re-cloned into the pOPIN vectors with the intention of testing expression in bacterial, insect, and if necessary mammalian cell lines in attempts to improve the protein solubility. The Vanderbilt plasmid maps are in Appendix I-f and pOPIN vector details are in Appendix I-g.

pSV281 (Vanderbilt University Structural Biology Center). This plasmid contains a TEV protease cleavable N-terminal 6His-tag, a kanamycin resistance gene and is under the control of the T7 promoter region.

pBG100 (Vanderbilt University Structural Biology Center). This plasmid contains a 3C protease cleavable N-terminal 6His-tag, a kanamycin resistance gene and is under the control of the T7 promoter region.

pBG101 (Vanderbilt University Structural Biology Center). This plasmid contains a 3C protease cleavable N-terminal 6His-GST-tag, a kanamycin resistance gene and is under the control of the T7 promoter region.

pOPIN-E (OPPF). This plasmid contains a non-cleavable C-terminal 6His-tag, an ampicillin resistance gene and is under the control of the T7 promoter region.

pOPIN-F. This plasmid contains a 3C-cleavable N-terminal 6His-tag, an ampicillin resistance gene and is under the control of the T7 promoter region.

pOPIN-J. This plasmid contains a 3C-cleavable N-terminal 6His-GST-tag, an ampicillin resistance gene and is under the control of the T7 promoter region.

2.2.3 Cell strains

E. coli *strains used for cloning*. Mach1 (T1R) chemically competent cells (Invitrogen) were used when cloning full-length TEAD and TAZ cDNA plasmids. XL1-blue subcloning-grade chemically competent cells (Stratagene) and DH5α cloning efficiency chemically competent cells (Invitrogen) were used for cloning procedures carried out with constructs in Vanderbilt vectors and TEAD4 pOPIN constructs. Omnimax (2 T1 R) chemically competent cells were used for cloning procedures carried out at OPPF.

Expression E. coli *strains*. BL21 (DE3), Arctic Express (DE3)^{TM} RIL/RP, and Rosetta (DE3)^{TM} (Stratagene) chemically competent cells were used for protein expression. Rosetta (DE3) cells were grown in the presence of 34 μg/ml chloramphenicol in addition to antibiotics required for selection of the expression vector.

2.2.4 Cell growth medium

Luria-Bertani (LB) medium. LB broth containing 10 g Tryptone, 5 g Yeast extract, and 10 g NaCl per 1 L distilled H_2O (Milli-Q water purification system, Millipore) was sterilised by autoclave and used to grow all strains of *E. coli*. Stock solutions of antibiotic (100 mg/ml ampicillin or 50 mg/ml kanamycin) were added to the medium using a dilution factor of 1/1000 prior to culturing the cells. LB agar was made according to the manufacturers instructions using distilled water and autoclaved. The agar was allowed to cool enough to be handled, the antibiotic was then added and the medium poured into 90mm Petri dishes. LB agar plates were stored at 4°C.

M9 minimal medium. M9 medium was used to grow cells producing labelled proteins for use in NMR spectroscopy. 1 L 10x M9 stock solution containing 67.82 g Na_2HPO_4 (anhydrous), 30 g KH_2PO_4, 5 g NaCl was prepared and autoclaved. Each 1 L labelled M9 medium contained 5 g D-glucose or 2 g ^{13}C-labelled D-glucose, 1 g ^{15}N-labelled NH_4Cl, 100ml 10x M9 stock, 1 ml 1 M $MgSO_4$, 50 µl $CaCl_2$, 650 µl trace elements stock solution (Appendix III-a), and 1 ml MEM Vitamin solution, and was made up to volume with autoclaved distilled H_2O in a sterile measuring cylinder. The medium was then transferred into a sterile 5 L conical flask (> 1 L volume) or 2 L conical flask (0.5-0.75 L volume).

2.3 Cloning

2.3.1 Polymerase Chain Reaction (PCR)

An initial temperature screen was carried out to determine the optimal annealing temperature for each reaction. A temperature gradient was set up for the annealing step of the PCR using an MJ Research PTC-200 Thermocycler. The final volume for each reaction was 20-25 µl. Successful reactions were scaled up using the best annealing temperature with reaction volumes of > 50 µl. Mini-prepped plasmids (approx. 20 ng/µl) being used as template DNA were diluted 1/100 in sterile water. PCR products from the first round of nested PCR being used as template DNA were not diluted. After transformation of competent cells with the insert-containing plasmid (section 2.3.4),

colonies were cultured in 200 μl LB medium for at least 5 hours and PCR was used to identify positive clones using the cells as template DNA.

KOD polymerase. This enzyme was used to amplify large DNA fragments (> 1 kb) for the production of cDNA and some TEAD domains. A typical reaction mixture is shown in Table 2.2. The temperature cycle for this enzyme is in Table 2.3.

Table 2.2 Components of a 25 μl PCR reaction when using KOD polymerase

Volume (μl)	Reagent
2.5	10x buffer (supplied with KOD)
1.5	25 mM $MgSO_4$ (supplied with KOD)
2.5	2 mM dNTPs (supplied with KOD)
0.75	Each primer (25 μM working stock)
1.0	Template DNA
0.4	KOD polymerase (1 U/μl)
15.6	Sterile H_2O

Table 2.3 Temperature cycle for KOD PCR

Step	Temperature	Time
1. Initialisation	95°C	1-2 min
2. Denaturation	95°C	20-30 sec
3. Annealing	Range/optimum °C	30 sec
4. Elongation	68°C	45-60 sec
5. Amplification	Go to step 2	36-38x
6. Final elongation	68°C	10 min
7. End	4°C	-

Phusion polymerase. This high fidelity enzyme was used to amplify the majority of DNA fragments. A typical reaction mixture is shown in Table 2.4. The temperature cycle for this enzyme is in Table 2.5.

Table 2.4 Components of a 25 μl PCR reaction when using Phusion polymerase

Volume (μl)	Reagent
5	5x buffer (HF1B)
0.5	10 mM dNTPs (10 μl each stock dNTP + 60 μl H_2O)
2	Each primer (10 μM working stock)
0.3	Template DNA
0.25	Phusion polymerase (2 U/μl)
18	Sterile H_2O

Table 2.5 Temperature cycle for Phusion PCR

Step	Temperature	Time
1. Initialisation	98°C	30 sec
2. Denaturation	98°C	5 sec
3. Annealing	Range/optimum °C	15 sec
4. Elongation	72°C	25 sec
5. Amplification	Go to step 2	35x
6. Final elongation	72°C	10 min
7. End	4°C	-

Taq polymerase. Taq was used to screen transformed cells for the plasmid containing the DNA insert. A longer initial step was used to ensure complete cell lysis in order for the enzyme to gain access to the DNA. Table 2.6 shows the components of a typical reaction and Table 2.7 illustrates a typical temperature cycle.

Table 2.6 Components of a 10 μl PCR reaction when using Taq polymerase

Volume (μl)	Reagent
1	10x buffer (HF1B)
0.2	10 mM dNTPs (10 μl each stock dNTP + 60 μl H_2O)
1	Each primer (10 μM working stock)
1	Template – colony cells
0.1	Taq polymerase (5 U/μl)
5.7	Sterile H_2O

Table 2.7 Temperature cycle for Taq cell screen PCR

Step	Temperature	Time
1. Initialisation	95°C	5 min
2. Denaturation	95°C	5-30 sec
3. Annealing	Range/optimum °C	15-30 sec
4. Elongation	68-72°C	35-90 sec
5. Amplification	Go to step 2	30-38x
6. Final elongation	68-72°C	10 min
7. End	4°C	-

2.3.2 DNA analysis and purification

Agarose gel electrophoresis. Agarose gels of 30 ml or 50 ml containing 1% agarose, 1X TAE buffer (Appendix III-h) and a 1/20000 dilution of a 10 mg/ml ethidium bromide solution were used to visualise PCR products. 5 µl PCR product was added to 1 µl 6x buffer and electrophoresed using a Biorad gel system at a constant voltage of 110V. For size determination, the DNA ladders (NEB) shown in Figure 2.1 were used. Gels were visualised using a Toyobo FAS-I electronic UV transilluminator.

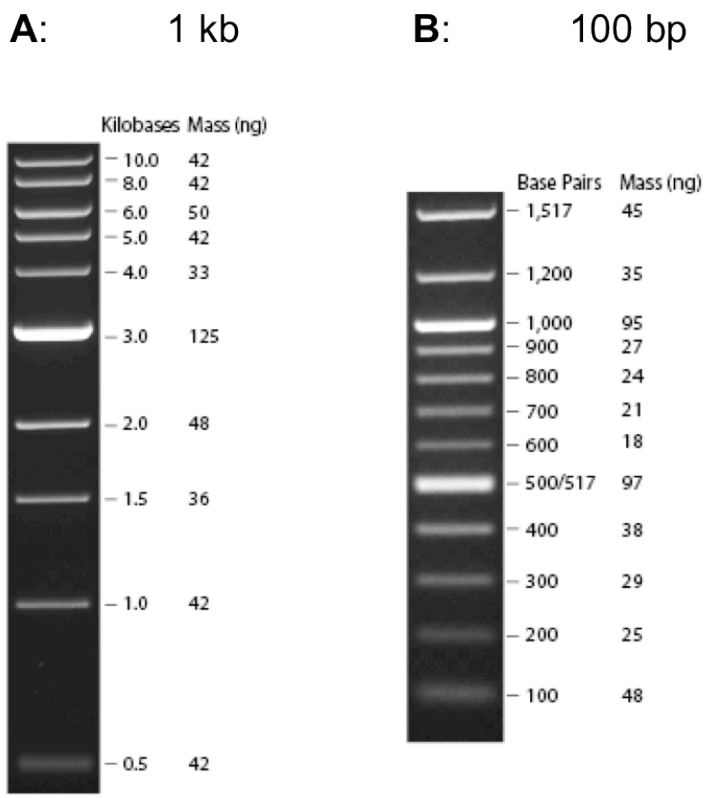


Figure 2.1 Markers used for DNA fragment size determination. (A) NEB 1 kb DNA ladder and (B) NEB 100 bp DNA ladder.

DNA purification. Nucleic acid purification columns for use with a micro-centrifuge (Eppendorf) were purchased as part of various DNA purification kits (PCR/Gel purification or miniprep kits,Table 2.1). For PCR products that produced a single band on an agarose gel, the PCR product was used directly in the PCR purification kits. For those that produced more than one band, the whole PCR reaction mixture was electrophoresed, cut out of an agarose gel and then purified. Whole plasmids were purified using miniprep kits. Phosphorylated full-length TEAD and TAZ DNA fragments were purified using phenol-chloroform extraction.

Phenol-chloroform extraction. All steps were carried out at 4°C. A 10 µl phosphorylation reaction mixture was made up to 50 µl with sterile H_2O, 50 µl phenol-chloroform (lower phase) was added to the tube and mixed by inverting the tube several times. The tube was centrifuged at 10,000 rpm for 5 min. The top layer was transferred to a fresh tube containing 50 µl chloroform and mixed. The tube was centrifuged at 10,000 rpm for another 5 min. The top layer was transferred to a fresh tube and subjected to ethanol precipitation.

Ethanol precipitation. 100 µl cold 100% ethanol and 5 µl 3 M sodium acetate pH 5.2 were added to 50 µl DNA and frozen at -80°C for at least 15 min. After warming the samples (hand temperature) enough to reduce the viscosity, the tube was centrifuged at 10,000-13,000 rpm at 4°C for 20-30 min. The supernatant was carefully pipetted off and 150 µl 70% ethanol was added to the DNA pellet. The tube was centrifuged at 4°C for 20-30 min and the supernatant discarded. The tube was left at room temperature for up to an hour or at 4°C overnight to allow the residual ethanol to evaporate. The DNA pellet was resuspended in 10 µl TE buffer.

2.3.3 Insertion of DNA fragments into plasmids

Vanderbilt vectors. DNA fragments and vectors were digested with restriction endonucleases in the supplied recommended buffer (a double digest reaction was preferentially carried out when buffers were compatible). Typically, 1.5-2 µl enzyme (20 U/µl) and 50 µl DNA (~20 ng/µl) were used in each reaction. High-fidelity restriction endonucleases were incubated with DNA at 37°C for 2-3 hours, with the exception of EcoRI which had a total incubation time of 1 hour. Fast-digest enzymes

were incubated with DNA for approximately 1 hour. PCR purification kits were used to terminate the reactions.

DNA fragments were ligated into the desired vector using T4 DNA ligase with the supplied buffer. A 20 μl ligation reaction typically contained 10 μl insert DNA (~10-20 ng/μl), 2-5 μl vector DNA (20 ng/μl) and 1 μl ligase. Reactions were carried out overnight in PCR tubes at 16°C.

pBluescript SK-. Blunt ended full-length (fl) DNA fragments (5 μl) were phosphorylated by 1 μl T4 PNK β in a 10 μl reaction containing ATP at 37°C for 30 min followed by inactivation at 65°C for 10 min. The phosphorylated DNA was purified by phenol-chloroform extraction and ethanol precipitation prior to ligation. Ligation mixtures containing 5 μl phosphorylated DNA insert, 1 μl dephosphorylated EcoRV digested pBluescript SK- vector (50 ng/μl) and 1.5 μl T4 DNA ligase with the supplied buffer, were incubated for 3 days (over the weekend) at 18°C.

pOPIN vectors. DNA fragments were inserted into HindIII- or KpnI-cleaved pOPIN vectors by ligation independent cloning (see Chapter 3 for details).

2.3.4 Transformation of E. coli *cells*

To insert the plasmids into cloning cell strains, 7.5-10 μl ligation mixture was added to 10 μl cells on ice and left for 30 minutes. The cells were heat-shocked at 45°C for 45 sec and returned to ice for 5 min. 300-600 μl room temperature LB medium (or for full-length TEAD and TAZ plasmids, SOC medium) was added to the cells which were immediately incubated at 37°C, shaking for 1-2 hours. At least 100 μl transformation mixture was spread onto an LB agar plate containing the appropriate antibiotic.

When a pure plasmid was used for transformation into expression cell strains, 1 μl plasmid was added to 10 μl chemically competent cells. The rest of the procedure was carried out as described above.

2.3.5 Sequencing

GATC biotech carried out sequencing of most plasmids. More recently Source Bioscience were used to sequence the plasmids (Sav). DNA sequences were translated into protein sequences using the Expasy translate programme and aligned with expected sequences using the EBI ClustalW or Align programmes.

2.4 Downstream protocols

2.4.1 SDS-PAGE analysis

All protein samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) using 15% bis-acrylamide gels in the Novex gel system (Invitrogen). The samples were prepared by mixing with 2X loading buffer (Appendix III-c) and boiling for 5-10 min. 5-15 μl samples were loaded onto the gel and electrophoresed in a 1X Tris-glycine buffer containing SDS (10X solution is shown in Appendix III-d), using a current of 60-100 mA. Gels were incubated with Coomassie Blue stain (Appendix III-e) for 30 min after being heated in a microwave for 30 sec. Gels were subsequently de-stained (Appendix III-f) and dried (Appendix III-g) in cellophane (Invitrogen). Unless otherwise stated, the Fermentas 'Unstained Protein Molecular Weight Marker' (Figure 2.2) was used to determine the size of proteins.

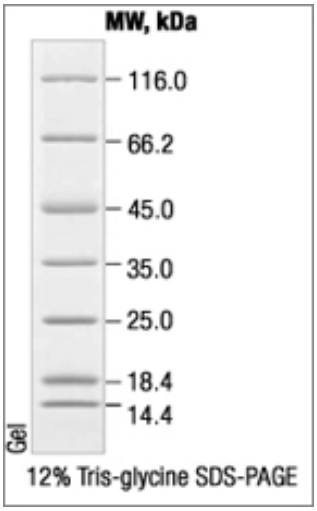


Figure 2.2 Fermentas 'Unstained Protein Molecular Weight Marker' used during SDS-PAGE. The marker was supplied in loading buffer; typically 5 μl marker was loaded into the well.

2.4.2 Protein expression and purification

Initial expression trials. Various domains and combinations of domains from the Hippo pathway proteins were initially transformed into chemically competent BL21(DE3) cells using standard protocols (section 2.3.4) and spread onto LB agar plates containing the appropriate antibiotic (50 μg/ml Kanamycin for Vanderbilt vectors or 100 μg/ml ampicillin for pOPIN vectors). Colonies were transferred into 5-10 ml LB

broth with antibiotic and grown to an OD_{600} of 0.4-0.6 at which point a glycerol stock was made (0.5 ml 50% glycerol and 1 ml culture, frozen on dry ice and stored at -80°C). At an OD_{600} of 0.6-0.9 the cells were induced to express the protein by way of the T7 promoter using 0.5 mM IPTG at 37°C for up to three hours. 1 ml samples were taken pre- and post-induction. The cell samples were centrifuged and resuspended in 100 µl water and 100 µl 2X SDS-PAGE buffer. Expression was assessed by SDS-PAGE.

Purification trials. Primary cultures of 1-10 ml (LB and antibiotic) were, unless otherwise stated, started from glycerol stocks and grown overnight, shaking at 180-220 rpm at 37°C before being sub-cultured in volumes of between 100 ml and 1 L. Cells were induced as previously described. After induction, cells were centrifuged at 3000-6000 g and resuspended in His buffer A (Appendix III-i). Cell lysis was carried out on ice using a Branson digital sonicator for a total time of 30-60 sec (0.5 sec pulses with 3 sec between bursts – saved as programme 1) at 60-80% amplitude. The lysate was centrifuged at high speed (16000-65000 g) and the supernatant retained. Proteins were purified using an ÄKTA Purifier 10 chromatography unit (GE Healthcare) by passing the cell lysate through a pre-packed nickel affinity column (HisTrap FF column; GE Healthcare). After washing the column with 3-5 column volumes of His buffer A, the bound protein was eluted with His Buffer B (Appendix III-i) in an imidazole gradient up to a concentration of 500mM.

Expression of labelled protein for use in NMR experiments. Primary cultures of 25-50 ml LB containing antibiotic were grown as described for purification trials. The cells were separated from the LB medium by centrifugation. M9 medium supplemented with ^{15}N-labelled NH_4Cl or ^{15}N-labelled NH_4Cl and ^{13}C-labelled glucose was used to resuspend the cells, which were then grown in 0.5-2 L M9 medium at 37°C. Subsequent induction and protein purifications were carried out as for unlabelled protein.

2.4.3 NMR spectroscopy

Sample preparation. Proteins were exchanged into NMR buffers by concentration and dilution using disposable centrifugal concentrators (Amicon/Sartorius/Millipore) with MW cut-offs at least 2-3 times lower than the MW of the protein, or using an Amicon stir cell concentrator with a cellulose ultra-filtration membrane (3 kDa cut-off). 5-10% D_2O was added to each sample, which was centrifuged to remove precipitate and

then the sample was transferred to a standard NMR tube or Shigemi tube depending on the amount of protein.

Experimental details. Nuclear magnetic resonance (NMR) experiments were performed using a Varian Unity INOVA spectrometer operating at the nominal proton frequency of 600 MHz, employing a room temperature triple resonance 5 mm probe equipped with pulse field gradients (PFG) along the z-axis. Unless otherwise stated, all datasets were acquired using Varian BioPack pulse sequences. Spectra were processed using NMRPipe/NMRDraw (Delaglio et al. 1995) and analysed with CCPN Analysis (Vranken et al. 2005). ^{1}H and ^{15}N chemical shifts were referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) (Wishart et al. 1995).

CHAPTER 3: TEAD PROTEINS AND THEIR INTERACTION WITH YAP

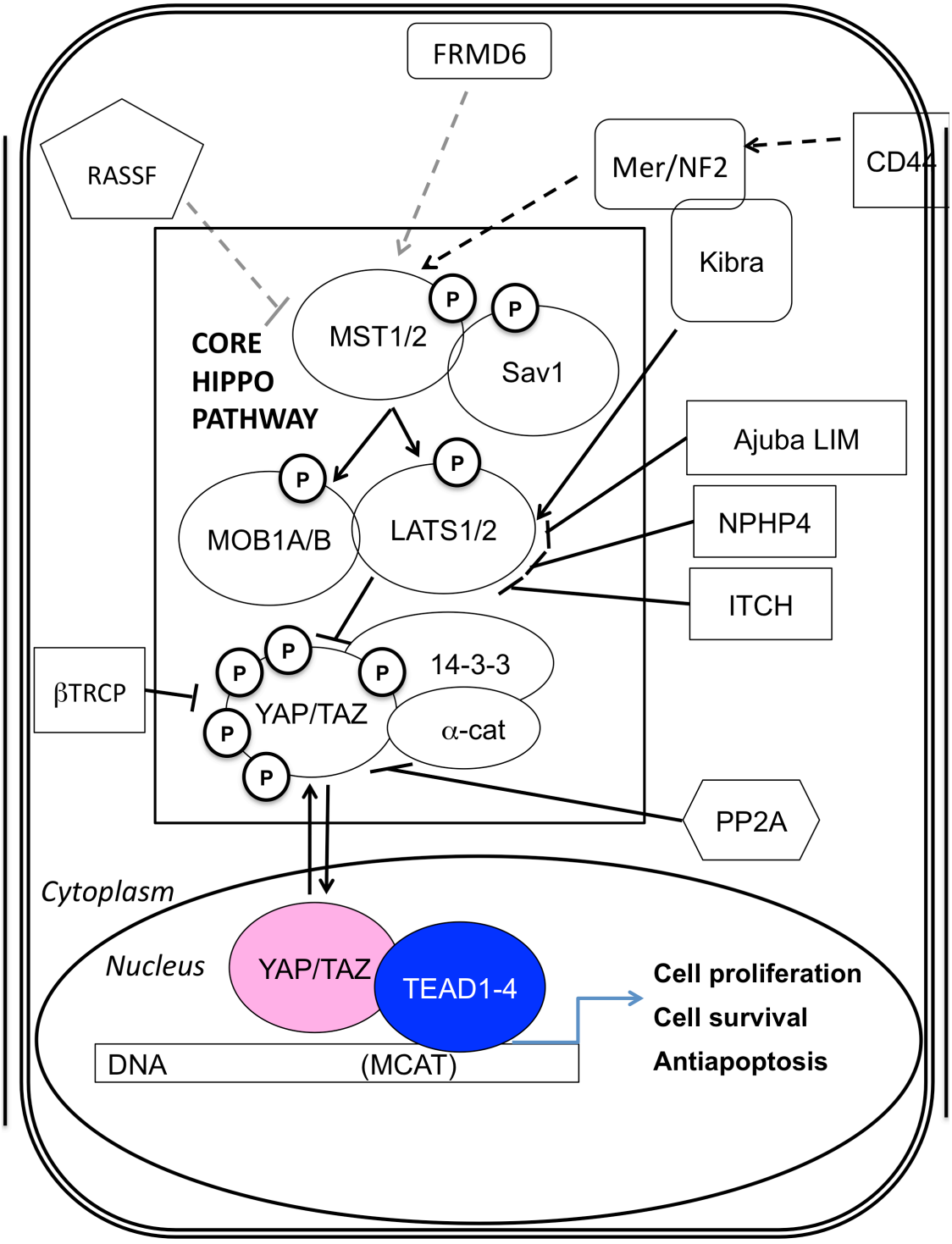


Figure 3.1 TEAD (blue) and YAP (pink) are located within the nucleus when the Hippo pathway is inactive. The two proteins interact to form a transcription factor complex that leads to the up-regulation of target genes.

3.1 Introduction

TEAD transcription factors bind to the MCAT DNA motif (5'-TCATTCCT-3') and similar DNA sequences through an N-terminal tri-helical DNA-binding domain (DBD) (Figures 1.7 and 3.2); the solution structure of a modified form of the TEAD1 (TEF-1) DBD was published in 2006 (Anbanandam et al.). TEAD proteins are functional only in combination with co-activators, one of which is YAP. The Hippo pathway regulates YAP-TEAD interactions by YAP phosphorylation that subsequently leads to cytoplasmic sequestration (Figure 3.1). The DBD of TEAD contains a nuclear localisation sequence in the third α-helix that allows the YAP-TEAD complex access to the nucleus resulting in the transcription of target genes (Goulev et al. 2008).

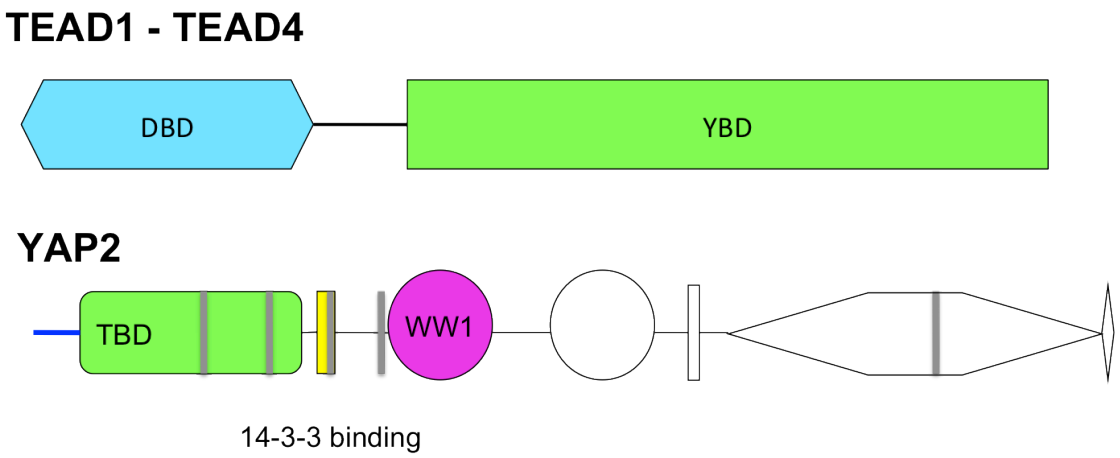


Figure 3.2 Representation of the TEAD and YAP domains studied in this chapter using the colour scheme seen in Figure 1.7. As for Figure 1.7, the domains and linker regions are to relative scales. Domains that do not feature in this chapter are not labelled or shown in colour.

The YAP-binding domain (YBD) of TEAD proteins (Figures 1.8 and 3.2) comprises the C-terminal region and is approximately 230 residues in length. The YBD binds the N-terminal TEAD binding domain (TBD) of YAP. The TBD is 107 amino acids long, however smaller fragments within this region are capable of binding TEAD with lower affinity than the full length TBD (Vassilev et al. 2001). The C-terminal region of YAP is thought to be a transcriptional co-activation domain and has no predicted secondary structural elements.

At the beginning of this project the functional interactions between TEAD (YBD) and YAP (TBD) were fairly well characterised but the structural details of the proteins were unknown. To really understand how a particular protein works and interacts with

binding partners, structural information at the atomic level is needed. Two common biophysical techniques that can yield such information are X-ray crystallography and NMR spectroscopy.

X-ray crystallography is used to determine a 3D structural model of proteins that are in a crystalline (i.e. solid) state. Proteins of almost any size can be studied using this technique but there are several limiting steps including crystallisation, diffraction and data analysis. NMR spectroscopy is used to study structural features and biochemical properties of proteins in solution. Traditionally, proteins should be 20 kDa or less for successful NMR experiments, though with the advent of more powerful spectrometers, better NMR pulse sequences (such as TROSY) and data processing/reconstruction, and more complex isotope labelling procedures, it is possible to study larger proteins. Both X-ray crystallography and NMR spectroscopy require pure, homogeneous, concentrated protein samples; therefore careful protein preparation is imperative for obtaining usable data.

The aim of the work described in this chapter was to use biophysical techniques to better understand the structural properties of the two discrete TEAD domains (DBD and YBD) individually and in complex with YAP, with a long-term goal of studying the DNA-TEAD-YAP interactions. Unfortunately setbacks were encountered during protein purification, crystallisation and data analysis. Three research articles containing crystal structures of TEAD YBD both alone and in complex with the TBD of YAP were published before this work could be completed (Chen et al. 2010; Li et al. 2010; Tian et al. 2010).

3.2 Materials and methods

General materials and methods are described in Chapter 2. This section describes the production and analysis of relevant individual samples in more detail.

3.2.1 Production of full-length template cDNA

TEAD1, TEAD2/3 and TEAD4 cDNA plasmid construction. Total RNA was extracted from two adult Hd-rR (southern inbred strain) medaka fish using TRIzol® reagent. The isolated RNA was treated with DNAse I and concentrated by phenol-chloroform extraction followed by ethanol precipitation as described in Chapter 2. A cDNA library was synthesised from the purified RNA using the TaKaRa PCR kit.

Nested PCR using KOD Hot Start polymerase with standard protocols was used to amplify blunt-ended TEAD cDNA as a single round of PCR did not yield sufficient quantities of DNA for cloning (i.e. PCR products were barely visible on agarose gels). The primers labelled 'outer' (Table 3.1) and the cDNA library as template were used for the first round of PCR. The primers labelled 'inner', and PCR products from the first round of PCR as template, were used for the second round of PCR that produced the TEAD cDNA used in subsequent procedures.

The full-length (fl) TEAD DNA fragments were phosphorylated, purified and ligated into the pBluescript SK- vector as described in Chapter 2. The resulting plasmids were transformed into Mach1 competent *E. coli* cells and spread onto LB agar plates containing 100 μg/ml ampicillin. PCR screens were performed on several colonies of each construct for selection of successful clones. Several clones for each TEAD plasmid were sequenced (GATC Biotech).

Table 3.1 Primers used for nested PCR amplification of full length TEAD cDNA

Construct	Outer forward	Inner forward	Inner reverse	Outer reverse
TEAD1	TAGCTTGGAGA GCCCGACT	AGAGCCCGACT GCCAAGAT	TGTTCATTCTTTA ACCAGTCTGTAG AT	GGTGTTCATTC TTTAACCAGTC TGT
TEAD2/3	CACCATTGCGTC CAACGA	GCGTCCAACGA GTGGAGT	TCCCTGGCAAGCT GTTAGTC	AGAAGCTCCC TGGCAAGC
TEAD4	AGAGGTTCCCTC TCCATCGT	CATCGTCCCACC ACGTCT	TCAGTCTTTGACC AGCCTGT	CACCAGTGCC TCTGTGTCAG

YAP cDNA. Dr Makoto Furutani-Seiki (MFS, University of Bath) generated Kyoto-Cab inbred medaka strain YAP cDNA prior to the commencement of our collaboration.

3.2.2 Cloning of TEAD and YAP DNA fragments

Cloning of TEAD DNA into Vanderbilt vectors. All TEAD DNA fragments were amplified by KOD Hot Start polymerase using standard protocols (Chapter 2). A medaka whole genome cDNA library (produced by Makoto Furutani-Seiki) was used as template for the TEAD2/3 (239-462) fragment with an annealing temperature of 54°C using the forward and reverse primers 'mTEAD3_V239f' and 'Med_TEAD2-rev' (Appendix I-a). The TEAD4fl (1) cDNA plasmid was used as template for TEAD 4 (222-470) with annealing temperatures of 54.6°C to 63.8°C, using forward and reverse

primers ‘TEAD4_Y222f’ and ‘TEAD4_D470r’ (Appendix I-a). Both of the aforementioned TEAD (YBD) DNA fragments as well as the pSV281 (6-His tag, TEV cleavable) and pBG101 (6-His GST tag, 3C cleavable) vectors underwent double digest reactions with EcoRI and XhoI high-fidelity restriction enzymes (Chapter 2).

Cloning of YAP DNA into Vanderbilt vectors. All YAP DNA fragments were amplified using Phusion DNA polymerase using standard protocols (Chapter 2). The YAP TBD-WW1 (1-168) construct was amplified using an optimal annealing temperature of 64.6°C using forward and reverse primers ‘YAP_M1f’ and ‘YAP_T168r’ (Appendix I-a). YAP DNA and vectors were subjected to double digest reactions using EcoRI and HindIII high-fidelity restriction enzymes (Chapter 2).

High-throughput ligation independent cloning of TEAD and YAP DNA into pOPIN vectors (OPPF). TEAD and YAP DNA fragments were amplified using KOD Hot Start DNA polymerase with primers designed for ligation independent cloning into pOPIN vectors (Appendix I-a). 1.5 μl DNA was cloned into 1 μl vector (pOPIN-E, pOPIN-F or pOPIN-J) using the In-Fusion™ cloning enzyme at 42°C for 30 minutes. 40 μl TE buffer was used to terminate the reaction (yielding a total volume of 50 μl). The mixture was then used to transform Omnimax cells following standard transformation protocols (Chapter 2), and the cells were grown in 24 well plates containing 1 ml LB agar, XGal, IPTG, and carbenicillin for positive selection. Positive colonies were picked, grown overnight in 1 ml LB broth at 37°C and screened by PCR. Plasmids containing the insert were mini-prepped and glycerol stocks were made in 96 well plates.

3.2.3 Expression and purification of TEAD and YAP (TBD) protein constructs

The initial purification trial protocol (Chapter 2) was used for the successful purification of YAP (TBD-WW1) in both pSV281 and pBG101 expression vectors. The initial expression and purification conditions were not successful for the TEAD protein constructs, which tended to be insoluble. The following strategies were used in the attempt to optimise TEAD purification conditions: altering (generally lowering) induction temperature and time to slow protein production and possibly aid folding; adding the detergent Triton-X100; changing expression cell strain including Arctic Express with rare codons and Rosetta for use with rare codon strains (see Appendix II for rare codon analysis); re-cloning with different boundaries to reduce the effect of

flexible termini and into alternative vectors for potential differences in expression and effects of different affinity/solubility tags; and co-expression with the chaperone protein pREP4-GroESL to aid protein folding (a kind gift from David Cowburn, then New York Structural Biology Centre, now at Albert Einstein College of Medicine). The TEAD constructs described below were the most successful samples i.e. the TEAD4 used for X-ray crystallography is different to the TEAD4 used in NMR experiments; this is because the Vanderbilt vector construct yielded a 2D NMR spectrum and the pOPIN construct yielded protein crystals.

Optimal expression and purification conditions of TEAD2/3 (YBD) for use in NMR experiments. BL21 (DE3) chemically competent cells (10 μl) were transformed with the TEAD2/3 (239-462) pSV281 plasmid and pREP4-GroESL plasmid (1 μl of each), a glycerol stock was made and the primary culture for the NMR sample was started from this. A 10 ml primary culture was grown overnight in LB broth at 25°C. The primary culture was transferred into 0.5 L LB medium and grown to OD_{600} of 1.8 at 37°C. The cells were centrifuged at low speed (< 3000 g) before being transferred to 1 L pre-warmed M9 minimal medium supplemented with ^{15}N NH_4Cl. 50 μg/ml kanamycin was present in all cultures. The cells were induced overnight at 12°C with 0.25 mM IPTG. The cells were resuspended in His buffer A with 4% Triton X-100, lysed by sonication and purified by affinity chromatography as described for initial expression and purification trials (Chapter 2).

Optimal expression and purification conditions of TEAD4 (YBD) used for NMR and trypsin digestion. Rosetta (DE3) competent cells (10 μl) were transformed with the TEAD4 (222-470) pSV281 plasmid (1 μl) and a glycerol stock was made from which all subsequent cultures were started (in the presence of 50 μg/ml kanamycin). Primary cultures were grown overnight at 37°C, secondary cultures were grown to an OD_{600} of 0.6-0.9 at 37°C in either LB medium or M9 minimal medium supplemented with ^{15}N NH_4Cl and induced overnight at 12°C with 0.5 mM IPTG. TEAD 4 (YBD) pSV281 was purified in the same way as TEAD2/3 (YBD) pSV281.

Optimal expression and purification conditions of TEAD4 (YBD) for crystallisation and X-ray crystallography. Chemically competent BL21 (DE3) cells were co-transformed with the TEAD4 (171-240) pOPIN-F plasmid (amp^R) and the pREP4-

GroESL plasmid (kan^R) and a glycerol stock was made from which all subsequent cultures were grown (in the presence of 100 μg/ml ampicillin and 50 μg/ml kanamycin). Expression and purification conditions of TEAD4 (YBD) pOPIN-F were as described for TEAD4 (YBD) pSV281.

The 6His-tag was cleaved from TEAD4 (YBD) pOPIN-F using 3C preScission protease in His-buffer A with 10% glycerol and 1 mM DTT, on a rocking platform at room temperature overnight. The protein was concentrated and buffer exchanged into 50 mM Tris pH 7.5 and 150 mM NaCl, using a disposable centrifugal concentrator with a MW cut-off of 10 kDa.

Optimal expression and purification conditions of YAP (TBD-WW1) for NMR spectroscopy. BL21 (DE3) competent cells (10 μl) were transformed with the YAP (1-168) pSV281 plasmid (1 μl) and a glycerol stock was made from which all subsequent cultures were started (with 50 μg/ml kanamycin). 50 ml primary cultures were grown at 37°C overnight, 0.5-1.5 L secondary cultures were grown to an OD_{600} of 0.6-0.9 at 37°C in M9 minimal medium supplemented with ^{15}N NH_4Cl. Protein expression was induced for 3 hours at 37°C with 0.5 mM IPTG. YAP (TBD-WW1) was purified as described for the TEAD proteins.

3.2.5 Trypsin digestion of TEAD4 (YBD)

TEAD4 (YBD) pSV281 was concentrated using a 10 kDa cut-off disposable centrifugal concentrator and buffer exchanged into trypsin cleavage buffer (50 mM Tris pH 8, 20 mM NaCl, 1 mM EDTA, 5 mM DTT) to a final volume of 50 μl at an approximate concentration of 6 mg/ml. 10 μg trypsin was added to the protein, half was incubated at 37°C and the remaining half was incubated at room temperature. 2 μl samples were taken at time points of 15, 30, 60 and 120 mins, 4 hours and overnight; the samples were frozen immediately. 2X SDS loading dye was added and the samples were analysed by SDS-PAGE.

3.2.6 NMR spectroscopy

Only 2D ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectra (described in the results section of this chapter) were obtained for the TEAD-YAP interaction domains. NMR experiments were carried out as described in Chapter 2. Details of experimental conditions and parameters relating to data for YAP-binding domains of TEAD and TBD-WW1 domains of YAP shown in Chapter 3.3 are in Table 3.2.

Table 3.2 NMR buffer conditions and parameters for HSQC experiments of TEAD and YAP interaction domains

Construct	**Buffer**	**Temperature**	**^{15}N SW**	**^{15}N carrier frequency**
TEAD2/3 (239-462) pSV281	50 mM Tris, pH 7.5	25°C	1519 Hz	119 ppm
TEAD4 (222-470) pSV281	20 mM KH_2PO_4, 200 mM NaCl, 1 mM DTT, pH 5.5	37°C	1519 Hz	119.5 ppm
YAP (1-168) pSV281	50 mM KH_2PO_4, 200 mM NaCl, 2 mM EDTA, pH 6.5	25°C	2000 Hz	118.5 ppm

3.2.7 Crystallisation and X-ray crystallography

Production of protein crystals. High-throughput 96 well crystallisation screens (Structure Screen I&II, Molecular Dimensions Ltd) were carried out on TEAD2/3 (YBD)-6His-tag, TEAD4 (YBD) and TEAD4 (YBD)-YAP (TBD-WW1) using an Art Robbins Phoenix nano-dispenser robot. One successful crystal hit occurred for TEAD4 (YBD) pOPIN-F without affinity tag, in 0.1 M sodium citrate pH 5.6, 20% v/v 2-propanol, and 20% w/v PEG 4000. This condition was used as a basis for optimisation, leading to the final buffer condition: 0.1 M sodium citrate pH 5.6, 2% v/v 2-propanol, 15% w/v PEG 4000.

X-ray diffraction experiments. Native crystals were frozen in the final buffer condition with 25% v/v glycerol. A data set was acquired at the Diamond Light Source synchrotron on beam IO2 with X-ray exposures of 0.25 sec at 100% intensity. The data were processed using HKL2000 software (Otwinowski et al. 1997). As no similar structures were available, molecular replacement could not be used to obtain the phase information that would allow the TEAD4 structure to be solved. A number of experimental techniques (as described below) were utilised when trying to solve the phase problem.

Selenomethionine labelling of TEAD4 was carried out by culturing the cells in M9 medium (Chapter 2) with 1 g NH_4Cl and 10 g D-glucose, supplemented with 50 mg seleno-l-methionine, in an attempt to prepare crystals for multi-wavelength anomalous diffraction (MAD) X-ray experiments.

In an attempt to prepare crystals for multiple isomorphous replacement (MIR) X-ray data collection, native crystals were soaked in the well buffer containing various concentrations (between 0.5 mM and 5 mM) of either mercury (II) chloride, phenylmercuric acetate or potassium tetrachloroplatinate for various amounts of time (the shortest being ~ 10 sec).

Longer wavelength data were collected on native crystals at the ID29 beam-line, ESRF, Grenoble, in an attempt to use native sulphur single-wavelength anomalous diffraction (SAD). The beam-line contains an adjustable aperture to 50 μm and a mini-kappa goniometer. 3-4 images were acquired at 6 keV, a resolution of 2 Å with 0.2 sec exposures. The MOSFLM programme (CCP4i) was used in the attempt to process, index, and predict the mosaicity of the data. The STAC_kappa programme was then used to orient the crystal to the best angle for obtaining a symmetrical diffraction pattern and another 3 images were acquired. These 3 images were processed and indexed as before. The RADDOSE programme (CCP4i) was used to estimate and optimise the best parameters for data collection (including resolution, starting angle, degrees of rotation, exposure time and number of images). Finally datasets were collected and attempts were made to process these data using XDS (Kabsch 2010).

Molecular replacement was attempted after the publication of the three TEAD (YBD) structures. The online programme Balbes (Long et al. 2008) was used to obtain an electron density map, and refinement was attempted using CCP4i (Potterton et al. 2003).

3.3 Results

3.3.1 TEAD cDNA isoforms

Medaka cDNA for full-length TEADs was cloned into the holding vector pBluescript SK- as PCR template for smaller TEAD domains. Several full-length (fl) TEAD cDNA plasmids (pBluescript SK-, amp^R) were sequenced, revealing the existence of different splice forms of medaka TEAD proteins. TEAD1 and TEAD2/3 splice form variants both contained an insertion in the DNA binding domain (DBD) (Figure 3.3). The DBD from human TEAD1 (a.k.a. TEF-1), which is 92% identical and 99% similar to the DBD of medaka TEAD1(s), comprises a bundle of three α-helices which bears close resemblance to the typical DNA binding homeodomain fold (Anbanandam et al. 2006). Secondary structure prediction using Jpred (Cole et al. 2008) identifies the anticipated three helices within the DBD of each non-variant TEAD protein. The insertion in TEAD1 and TEAD2/3 was not aligned with any known protein and so were not predicted to contain secondary structure.

TEAD4 was not found to have a DBD variant, however many of the TEAD4 plasmids contained a stop codon at the beginning of the YBD, in place of the codon for W78. It is possible that a TEAD4 protein consisting only of the DBD could be translated *in vivo*.

3.3.2 Protein expression and purification

Various domains and combinations of domains from TEADs and YAP were expressed in BL21 (DE3) *E. coli* cells. In some cases expression was poor upon induction or the protein was insoluble. In such cases alternative expression methods were utilised in an attempt to optimise the production of soluble protein, including the use of various expression strains of *E. coli*, co-expression with a plasmid containing the chaperone complex GroELS, and re-cloning fragments with altered boundaries and different expression vectors. Table 3.3 summarises the final expression and purification outcomes; all constructs were tested for purification potential but only the most successful constructs are described in detail. The numbering of TEAD4 was different for the pSV281 and pOPIN-F constructs because of differences in the predicted start of the sequence. The pSV281 primers were ordered first but cloning using genomic DNA was initially unsuccessful; subsequently the TEAD4 cDNA was successfully cloned and pOPIN primers were designed with the new numbering. Later the pSV281 TEAD4

(YBD) cloning was successful but the original numbers were used to identify the primers. Both TEAD4 constructs end at the C-terminal Asp.

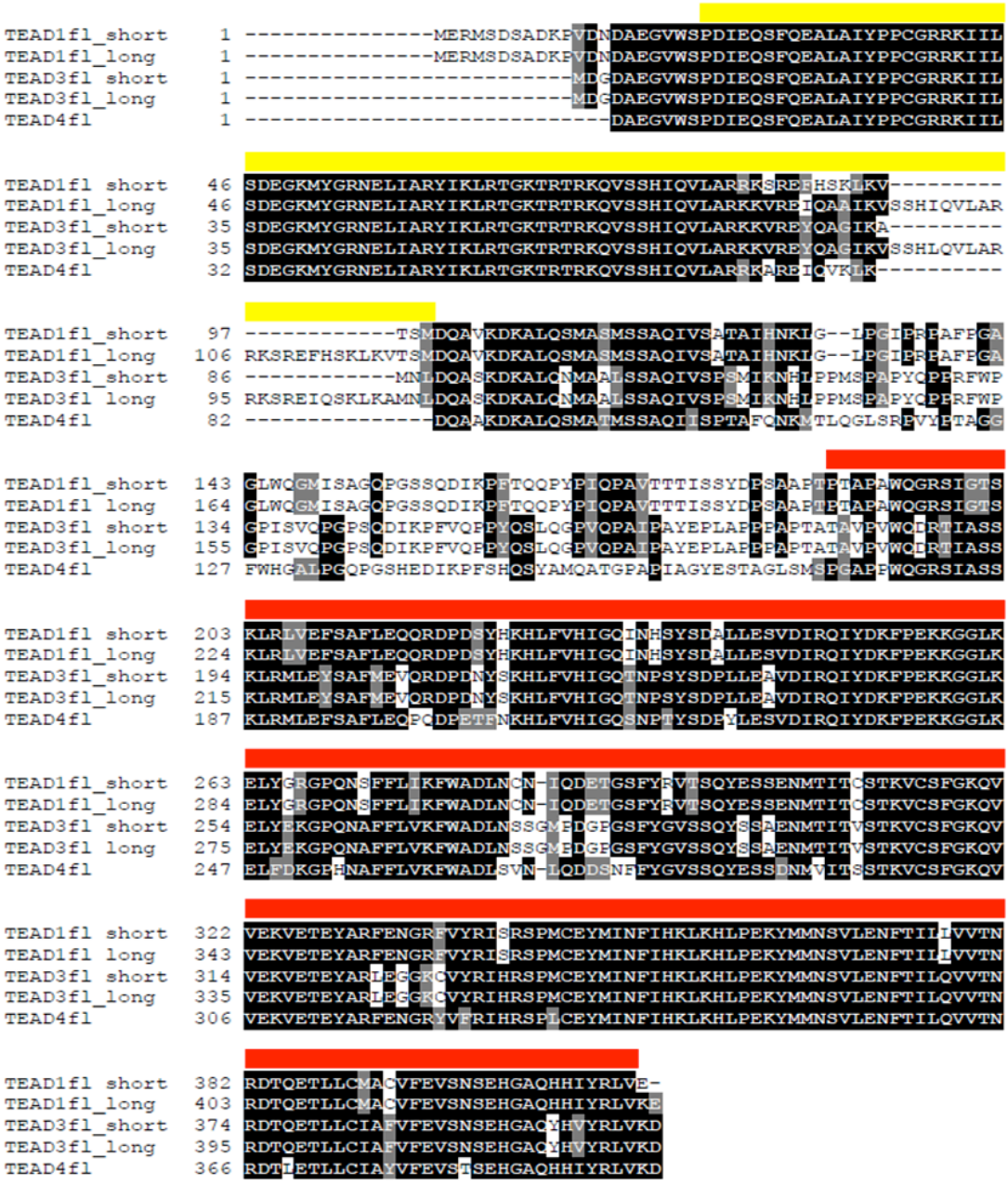


Figure 3.3 Full-length medaka TEAD sequence alignments from cDNA translation (Swiss Institure of Bioinformatics Expasy Translate programme). Alignments were made in ClustalW; TEAD1fl_short 84-96 and TEAD4 75-81 were manually moved to align with TEAD3fl_short. The shading was carried out using Boxshade; Black background represents identical alignments and grey backgrounds represent similar alignments. Yellow bars indicate the DNA binding domain (DBD), red bars indicate the YAP binding domain (YBD). The insertion in the DBD of TEAD1 and TEAD2/3 are at residue numbers 96-117 and 85-106.

Table 3.3 Summary of protein expression and purification outcomes for TEAD constructs and associated interaction domains of YAP and TAZ.

Protein	Domain	Expression	Solubility	Purification
TEAD 1	DBD	Good	Insoluble	**Unsuccessful**
TEAD 2/3	DBD	Good	Insoluble	**Unsuccessful**
TEAD 1	YBD	Poor	Insoluble	**Unsuccessful**
TEAD 2/3	YBD (239-462) pSV281	Good	20-50% soluble: Co-expressed with GroELS in BL21 (DE3) at 12°C with 0.25 mM IPTG	**Successful:** Nickel ion affinity chromatography
TEAD 4	YBD (222-470) pSV281	Good	40-60% soluble: Expressed in Rosetta (DE3) at 12°C with 0.5 mM IPTG	**Successful:** Nickel ion affinity chromatography
TEAD4	YBD (171-398) pOPIN-F	Good	50-80% soluble: Co-expressed with GroELS in BL21 (DE3) at 12°C with 0.5 mM IPTG	**Successful:** Nickel ion affinity chromatography
YAP	TBD	Good	Insoluble	**Unsuccessful**
YAP	TBD + WW1 (1-168) pSV281	Good	Soluble: Expressed in BL21 (DE3) at 37°C with 0.5 mM IPTG	**Successful:** Nickel ion affinity chromatography

Final yields of TEAD2/3 were low (~0.5-1 mg/L), TEAD4 yields were better (~3-6 mg/L), and YAP (1-168) yields were reasonable (~5-10 mg/L).

3.3.3 The YAP-binding domain of TEAD is globular and folded

Purification of TEAD2/3 and TEAD4 (YBD) proteins used in NMR and trypsin digestion experiments. The most successful NMR sample from the TEAD proteins was from TEAD2/3 (YBD) in pSV281 when co-expressed with pREP4-GroESL (Piserchio et al. 2009) in BL21 (DE3) cells. The TEAD4 (YBD) pSV281 construct expressed in Rosetta (DE3) also yielded a 2D ^{1}H-^{15}N-HSQC spectrum. The elution profiles are shown in Figure 3.4.

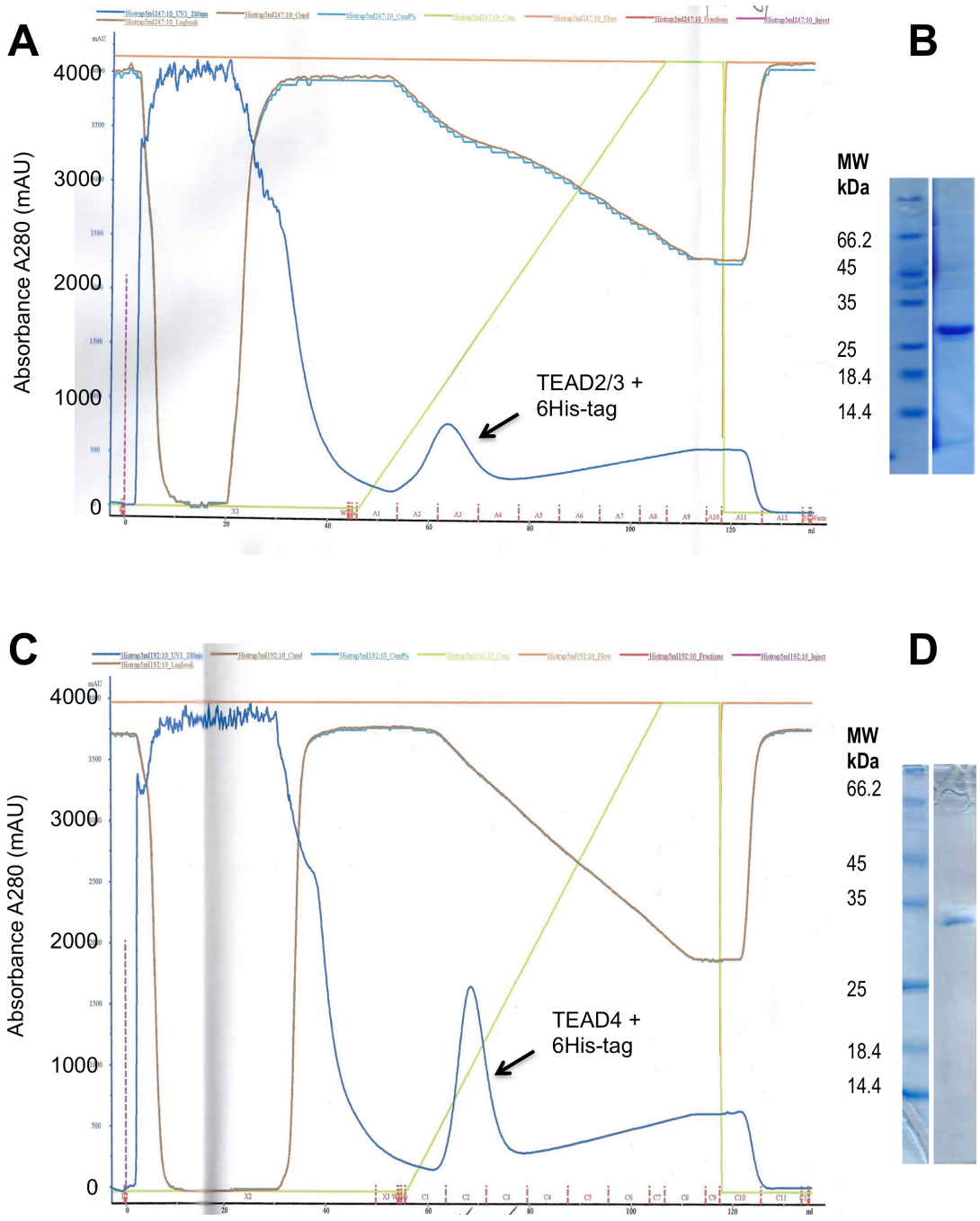


Figure 3.4 NMR sample purification and SDS-PAGE profiles for (A, B) TEAD 2/3 (YBD in pSV281) and (C, D) TEAD4 (YBD in pSV281). Volume (ml) and fractions are shown on the x-axis. The blue line in the elution profile signifies A_{280} and the green line represents the imidazole gradient. The proteins were purified using a 5 ml His-Trap FF column (GE Healthcare). (B, D) The purity of the proteins was verified using Coomassie blue stained SDS-PAGE gels; the marker is shown on the left hand side with the corresponding MW of each band annotated; purified TEAD protein, with an expected size of ~28 kDa, is shown on the right hand side of the gel images. 5 µl marker and 10 µl of the pooled peak fractions diluted 1:1 with sample buffer were loaded onto the gel.

Trypsin digestion of TEAD4 (YBD). The YAP-binding domain of TEAD4 was digested with trypsin to gain an indication of whether the domain contains exposed unstructured regions that would be susceptible to cleavage. A slight decrease in molecular weight was observed and a single distinct band was visible at around 26 kDa (Figure 3.5). This band represents a protein population that is resistant to protease digestion and is indicative of a globular domain.

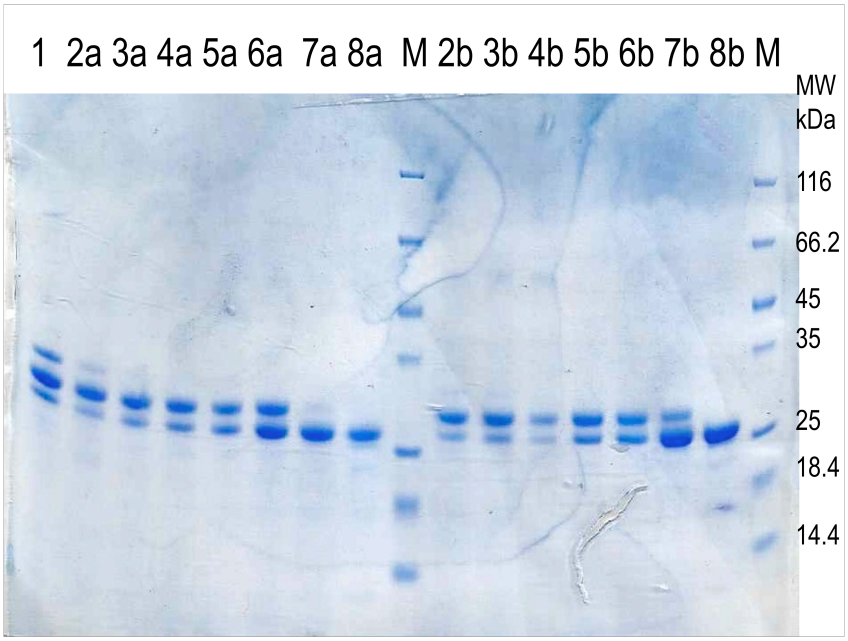


Figure 3.5 Trypsin digestion of TEAD4 (YBD), visualised on a Coomassie blue stained SDS-PAGE gel. Each lane represents a different incubation time of 0 (1), 8 min (2), 15 min (3), 30 min (4), 1 hr (5), 2 hrs (6), 4 hrs (7), and 20 hrs (8). Lanes labelled as 'a' and 'b' represent incubations at 37°C and room temperature, respectively.

NMR spectroscopy of TEAD2/3 and TEAD4. In order to verify that the TEAD transcription factor YBD is a structured domain as implied by trypsin digestion, ^{1}H-^{15}N-HSQC experiments were carried out on ^{15}N labelled TEAD2/3 and TEAD4. HSQC spectra provide protein 'fingerprints'. Each residue containing an NH within the backbone peptide bond is represented by a spot (peak) that comprises a signal (chemical shift) from the ^{1}H in one dimension (x-axis in this case) and the ^{15}N in the other dimension (y-axis). The chemical environment of the residue determines the location of each peak within the spectrum. The pattern and distribution of the peaks can provide information about protein structure and conformational stability.

The 2D ^{1}H-^{15}N-HSQC spectra (Figure 3.6) show roughly the expected number of backbone peaks i.e. number of residues minus proline and the N-terminal amino acid (233 expected for TEAD2/3 including His-tag residues, 217 expected for TEAD4). The peaks are well dispersed in both the ^{1}H and ^{15}N dimensions, indicating that secondary and tertiary structural elements are present. The TEAD2/3 (YBD) peak intensities are uniform, suggesting that the YBD adopts a single dominant conformation in solution. The TEAD4 (YBD) spectrum exhibits variable peak intensities, which may be indicative of conformational exchange or a population of misfolded/unfolded protein.

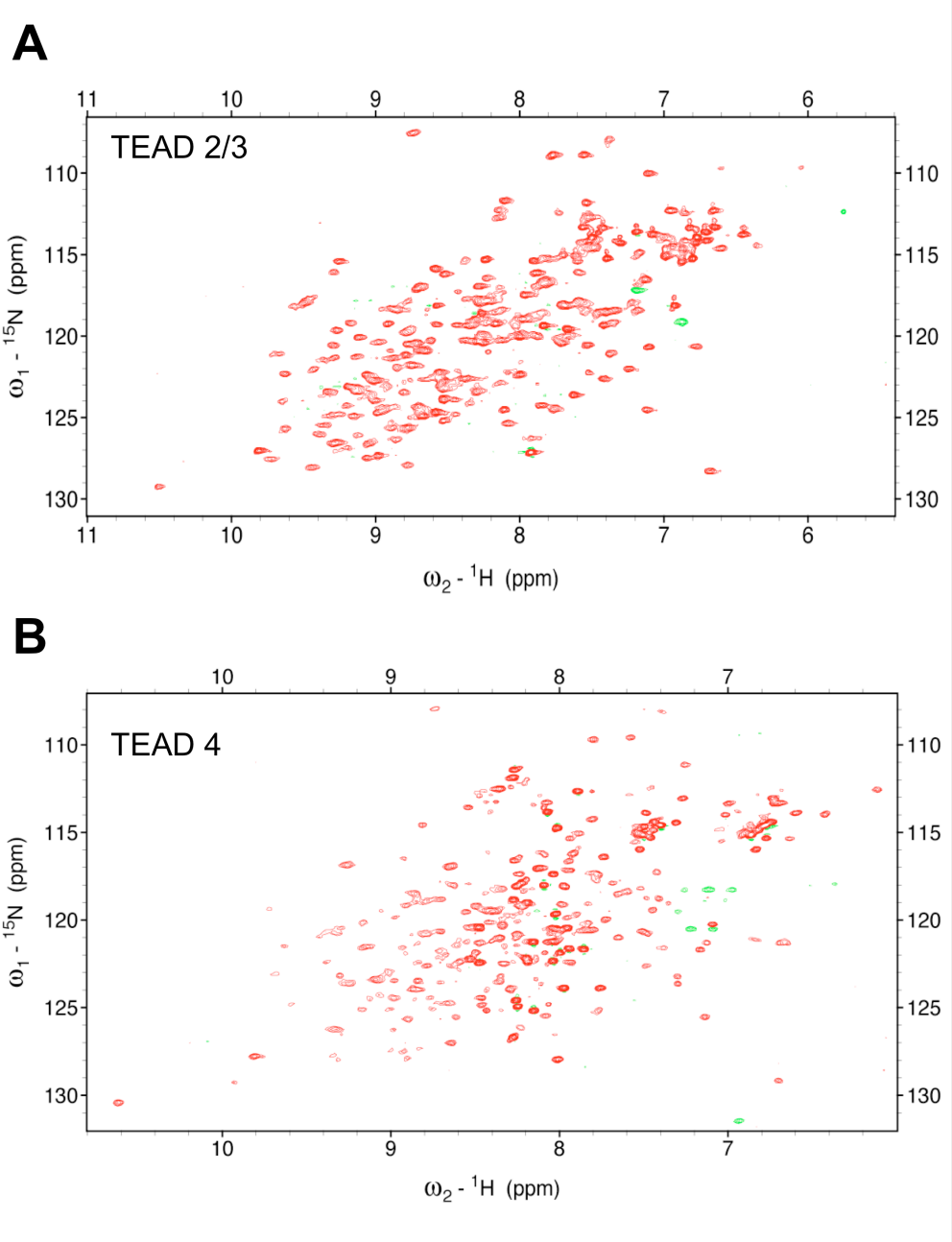


Figure 3.6 2D ^{1}H-^{15}N-HSQC spectra of (A) ^{15}N TEAD2/3 (YBD) acquired at 25°C with 256 scans and (B) ^{15}N TEAD4 (YBD) acquired at 37°C with 32 scans. Positive peaks are shown in red and negative peaks (arising from out of range folded positive peaks) are shown in green.

3.3.4 Purification and crystallisation of TEAD4 (YBD)

Overview of X-ray crystallography, data collection and processing. NMR experiments implied that the YBD of TEAD is folded, so we attempted to solve the structure, which at the time remained unknown, using X-ray crystallography. When proteins are crystallised and subjected to X-ray radiation diffraction patterns of regularly spaced spots (known as reflections) may be obtained; the X-ray scattering is caused by the electrons from each atom within the protein and variations in electron density cause variations in the intensity of each reflection. If a crystal diffracts to a high resolution (ideally at least 3Å) an atomic-level model of the 3D protein structure can, given the necessary parameters, be calculated. The protein crystal is rotated through 180° and a diffraction pattern collected at each rotation.

During data processing the diffractions must be indexed to identify the unit cell dimensions and space group, integrated into a single file, then merged to ensure only one of each reflection is present and scaled to make the intensity of each reflection consistent. To calculate an electron density map of the protein three pieces of information are needed: the wavelength of the X-ray (which is set in the experiment), the amplitude of the X-ray wave (which can be determined from the intensity of the reflection), and the phase of the X-ray wave (which cannot be directly measured as part of a single diffraction experiment). Thus, in order to calculate an electron density map on which to base the final structure, the 'phase problem' must first be solved. Various methods exist to overcome the phase problem, from using the phases from a similar structure to incorporating (or using naturally occurring) 'anomalous scatterers' e.g. heavy atoms, into the crystal and calculating the phase from the known structure of such atoms.

Crystallisation of TEAD4. The first limiting step of X-ray crystallography is producing protein crystals. Successful crystallisation was achieved with the YBD of TEAD4 (171-398) in pOPIN-F when co-expressed with pREP4-GroESL (Piserchio et al. 2009), purified using a HisTrap-FF column, and the 6His-tag cleaved (Figures 3.7 and 3.8). The final crystallisation buffer consisted of 0.1 M sodium citrate pH 5.6, 2% isopropanol and 15% PEG4000. The initial 'hit' that this buffer was derived from contained 0.1 M sodium citrate pH 5.6, 20% isopropanol and 20% PEG4000. Other 'hits' occurred in the Molecular Dimensions Structure screen I+II (HT96) F11 (1.6 M

$(NH_4)_2SO_4$, 0.1 M MES pH6.5 and 10% dioxane) and C9 (0.2 M MgCl, 0.1 M Tris pH8.5 and 30% PEG4000).

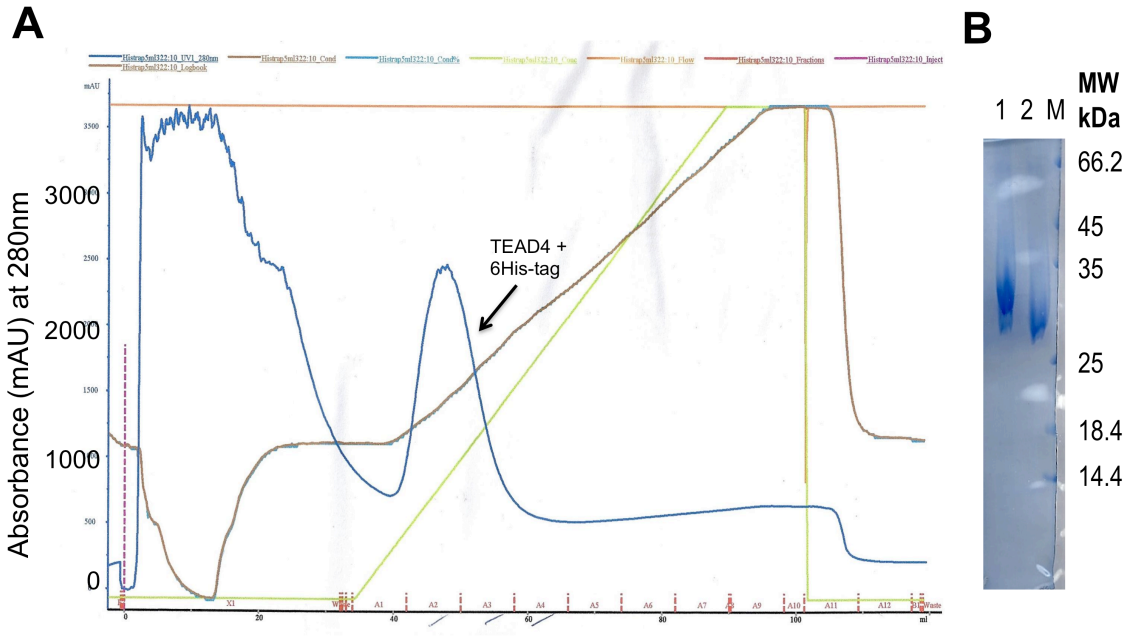


Figure 3.7 (A) Affinity purification elution profile from a 5 ml His-Trap FF column of TEAD4 (YBD) pOPIN-F used in successful crystal trials (A_{280} is represented by the blue line; imidazole gradient is shown in pale green). (B) The Coomassie blue stained SDS-PAGE gel shows 1 µl concentrated TEAD4 (YBD) with (1) and without (2) the His-tag, at ~ 28 and 26 kDa, used in crystallisation trials.

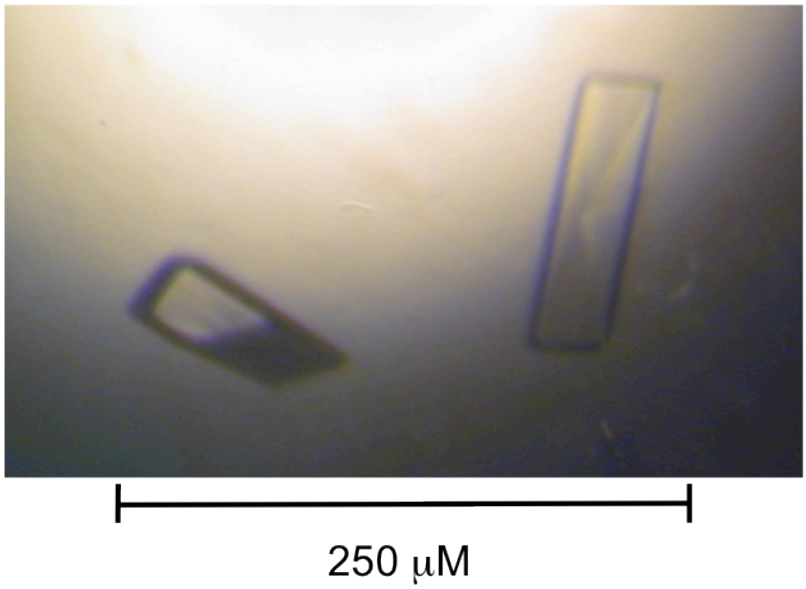


Figure 3.8 TEAD4 crystals at a final magnification of 500X in the optimal crystallisation condition 0.1 M sodium citrate pH 5.6, 2% isopropanol and 15% PEG4000. Crystals were approximately 125 µM in length.

The native crystals in Figure 3.8 diffracted to a resolution of 1.9 Å (Figure 3.9, Table 3.4) at the Diamond Light Source synchrotron beam (IO2) in Oxford, UK. At the time of data collection, there were no similar structures; hence an electron density map could not be calculated using phasing information from the molecular replacement method. In order to obtain difference data to solve the phase problem by MIR, the crystals were soaked in heavy atom-containing solutions, which either cracked the crystals, prevented diffraction or produced very mosaic diffraction datasets that were difficult to process and impossible to refine. Seleno-methionine labelled protein that could potentially be used for MAD did not crystallise. Longer wavelength X-rays were also utilised in the attempt to allow phasing by sulphur-SAD, but the data displayed signs of mosaicity and could not be properly processed. It was at this point that crystal twinning was first suggested (by Christoph Mueller-Dieckmann, ESRF, Grenoble).

Crystal twinning occurs when crystal domains are not in perfect alignment. There are various types of twin (as discussed in section 3.4) including macroscopic, merohedral and non-merohedral. Macroscopic twins can often (but not always) be seen under a light microscope whereas other types of twin are sub-microscopic and cannot be detected by eye. Some useful diagrams to illustrate the different types of twinning can be found in pages 388-394 of Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology (Rupp 2009), which cannot be reproduced here due to copyright.

Table 3.4 Native TEAD4 (YBD) crystal data processing statistics and parameters from data collected at the Diamond Light Source on beam IO2

Cryo buffer condition	25% glycerol added to: 0.1 M Sodium citrate pH 5.6, 2% isopropanol, 15% PEG4000	
Wavelength	0.976 Å	
Resolution Limit	1.86 Å	
Cell dimensions	62.9 Å, 89.5 Å, 109.2 Å, 90.00°, 90.00°, 90.00°	
Unit Cell	P 2 2 2_1	
Total reflections	602085	
Unique reflections	43964	
	50-1.98 Å	**2.05-1.98 Å**
R_{merge}	0.063	(0.403)
Completeness	96.4	(94.5)
I/σ(I)	22.1	(3.43)

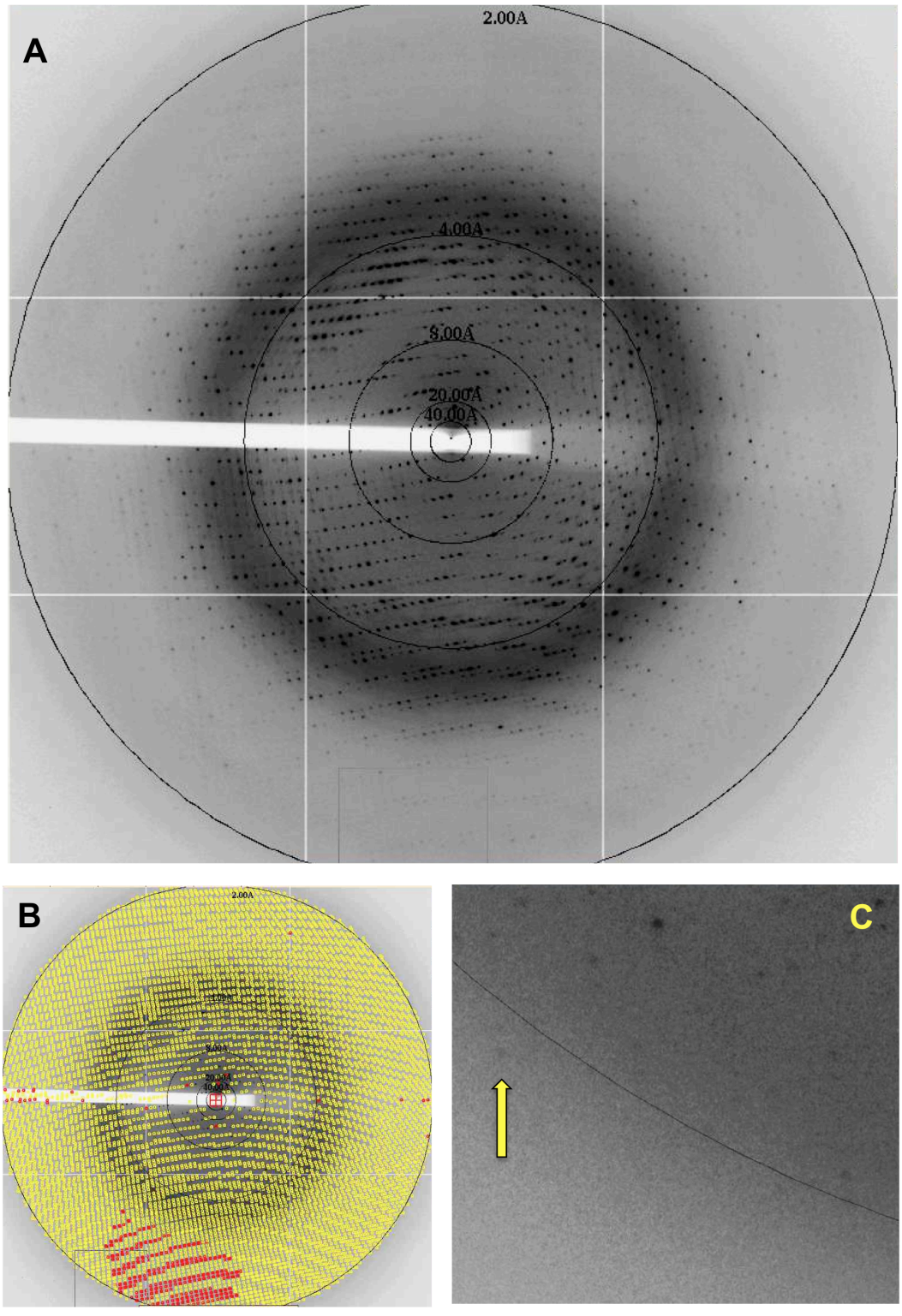


Figure 3.9 (A) Diffraction data for native TEAD4 (YBD) crystal. The data processing (using HKL2000 (Otwinowski et al. 1997)) that placed the crystal in the orthorhombic $P222_1$ space group is shown in panel B where yellow spots represent partial measurements and red represents overlapped spots. Panel C illustrates a diffraction spot beyond the 2 Å marker.

Shortly after the phasing attempts, three papers containing various structures of TEAD YBDs were published (Chen et al. 2010; Li et al. 2010; Tian et al. 2010). Electron density maps were calculated using these structures to model the phases by molecular replacement. Unfortunately the R-factors could not be reduced during the refinement stage and we returned to the idea that the crystals may be twinned. The native dataset was converted to CNS format using CCP4i (Potterton et al. 2003) and submitted to the UCLA 'Merohedral Twin Detector' programme, which is not affected by anisotropy, pseudo-centering or pseudo-non-crystallographic symmetry (Yeates 1997). The output graph indicates that the crystals were partially twinned (Figure 3.10). The difficulty experienced in processing the data at ESRF was also a sign of non-merohedral twinning. As perfect merohedral twinning did not occur, the twinning factor could not be corrected for and the structure of medaka TEAD4 YBD remains unsolved.

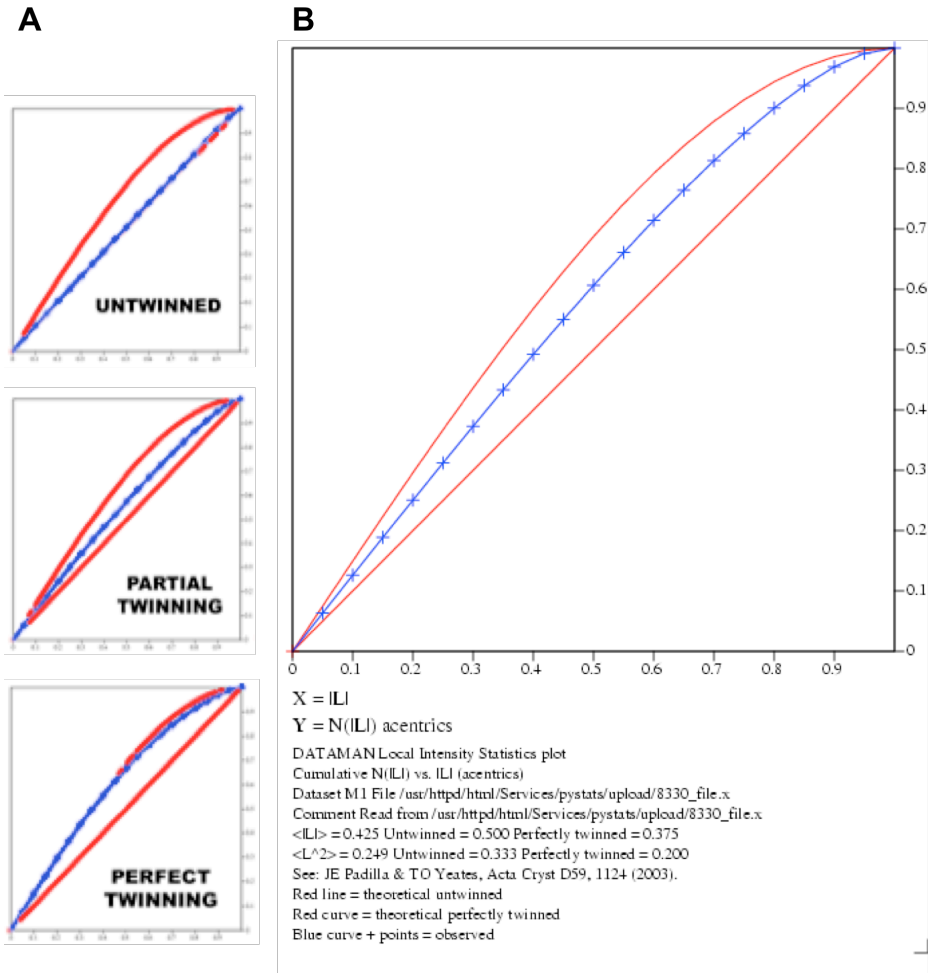


Figure 3.10 (A) Reference and (B) native TEAD4 crystal diffraction data output graphs from the UCLA 'Merohedral Twin Detector' using the Padilla-Yeates algorithm (Yeates 1997). Graphs show the local intensity statistics curve, N(|L|) versus |L|, where L = [I(A) - I(B)]/[I(A) + I(B)] and I(A) and I(B) are the intensities of nearby (local) unrelated reflections A and B.

3.3.5 YAP (TBD) protein purification and NMR spectroscopy

The TEAD binding domain of YAP expressed well in BL21 (DE3) cells but was insoluble and refolding attempts were unsuccessful. For the NMR sample, ^{15}N labelled YAP (TBD-WW1) in pSV281 was used. The ^{15}N labelled protein was expressed in BL21 (DE3) cells for 3 hours at 37°C and purified using standard protocols described in Chapter 2 (Figure 3.11).

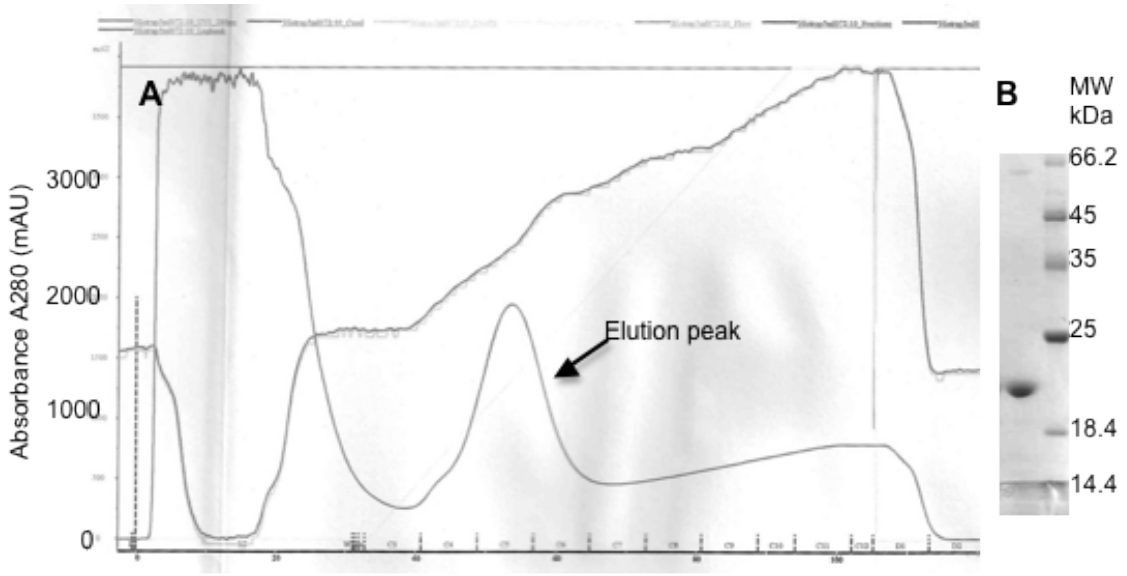


Figure 3.11 (A) Elution profile of ^{15}N YAP (TBD-WW1) from a 5 ml His-Trap FF column (GE Healthcare). (B) Coomassie blue stained SDS-PAGE analysis of purified ^{15}N-labelled YAP (TBD-WW1) with an expected MW of 20 kDa; 10 µl of the concentrated NMR sample diluted 1:10 with sample buffer (appendix III-c) (left) and 5 µl marker (right, with band MWs indicated) were loaded.

The 2D ^{1}H-^{15}N-HSQC (Figure 3.12) revealed that the N-terminal, TBD of YAP is natively unstructured. Most of the peaks are clustered in the centre of the spectrum, many of which overlap. The weak peaks that are more widely dispersed arise from the folded WW1 residues. In accordance with these results, Tian et al (2010) also found that the N-terminus of human YAP is natively unfolded.

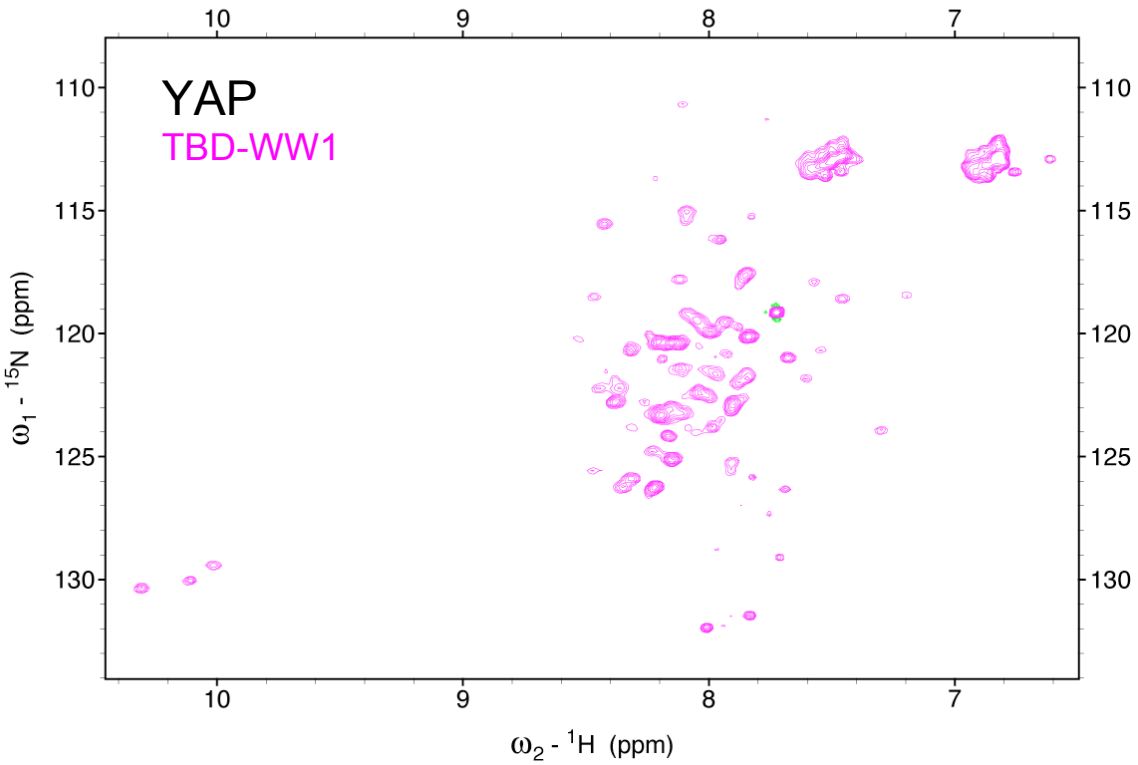


Figure 3.12 ^{1}H-^{15}N YAP (TBD-WW1) HSQC acquired at 25°C with 8 scans. Buffer conditions are shown in Table 3.2. The lack of peaks in the 1H 8.5-9.5 ppm region of the spectrum and the high proportion of overlapped peaks in the central region suggest that the protein lacks rigid secondary structure.

3.3.6 TEAD-YAP interaction

In order to verify that medaka fish YAP and TEAD proteins interact, and to set up co-crystal screens, TEAD4 (YBD) pOPIN-F was immobilised on a 5 ml nickel affinity column and used to pull-down pure YAP (TBD-WW1). SDS-PAGE analysis of the elution peak revealed the presence of both proteins (Figure 3.13) indicating that, as expected, they are capable of forming a complex.

Crystal screens were carried out and although some 'hits' occurred, successful crystallisation conditions could not be replicated on a larger scale. Following the publication of the mouse TEAD4-YAP (Chen et al. 2010) and human TEAD1-YAP (Li et al. 2010) interaction domain complex crystal structures in 2010, the decision to focus on other Hippo pathway protein domains was taken (Chapters 4 and 5).

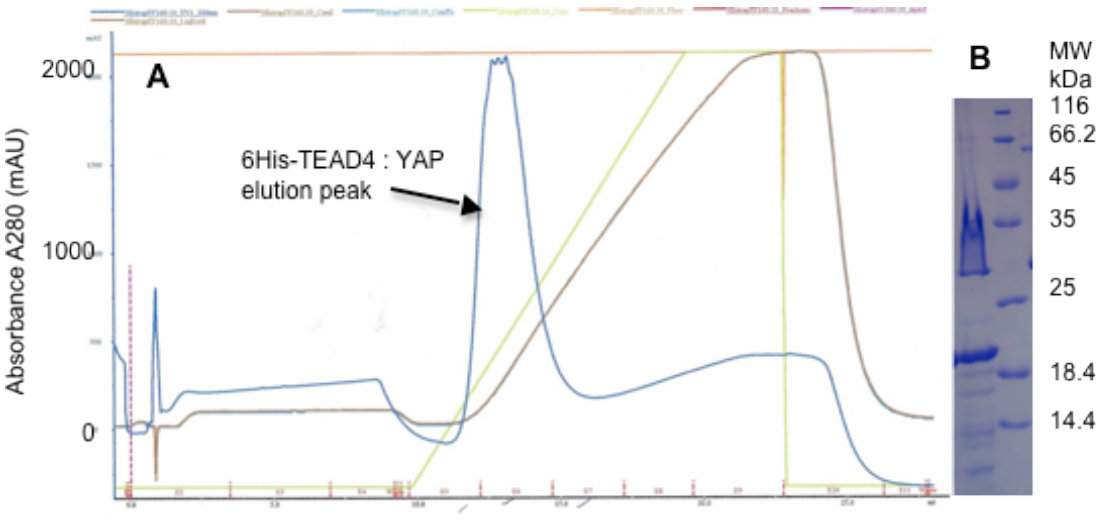


Figure 3.13 (A) Elution profile from a 5 ml His-Trap FF column, the imidazole gradient is shown in green and the A_{280} in blue. (B) Coomassie blue stained SDS-PAGE analysis of His-tagged TEAD4 (YBD) pull-down of cleaved pure YAP (TBD-WW1) on the left and the marker on the right (with band MWs indicated to the right of the gel). The band at approximately 28 kDa is TEAD4 and the band at 20 kDa is YAP. The affinity tag was removed from YAP prior to use in the experiment

3.4 Discussion

3.4.1 TEAD cDNA isoforms

Plasmids containing TEAD cDNA were constructed from medaka fish mRNA with the intention of using them as PCR templates. Sequencing the plasmids revealed variant full-length (fl) TEAD isoforms. TEAD1 (fl) and TEAD2/3 (fl) each had a primary structure close to the expected sequence (that was in the Ensembl database), they also both had a splice variant containing an insertion of around 20 amino acids within the DNA binding domain (DBD). It took two separate attempts to obtain a useful TEAD4 (fl) cDNA plasmid as all of the initial plasmids sequenced contained a stop codon in place of W78, at the N-terminus of the YBD.

The insertion of around 20 amino acids located within the DBD of TEAD1 is 68% identical and 84% similar to the predicted third α-helix (H3) (Cole et al. 2008) of the shorter non-variant TEAD1 (TEAD1(s)). The insertion in TEAD2/3 DBD has 68% identity and 90% similarity with predicted H3 of TEAD2/3(s). Therefore, although the lack of a corresponding inserted region in the database for alignment meant that the insertion was not predicted to be α-helical, it is likely that the insertion forms a fourth α-helix (an alignment of the four regions is shown in Figure 3.14a).

According to the original multiple sequence alignments of the full-length medaka TEAD proteins (Figure 3.3), the TEAD1(l) variant contains the insertion/theoretical helix between H2 and H3 (of TEAD1(s) non-variant), whereas in the TEAD2/3(l) variant the insertion/theoretical helix is apparently C-terminal to H3. This was the case even when the gap penalties were increased. When the DBDs of TEAD1 and 2/3 are aligned with human TEAD1 without the YBD (again using ClustalW2), the theoretical helix is C-terminal to H3 in both cases (Figure 3.14b), further highlighting the similarity between H3 and the insertion, which will now be called theoretical H4 (tH4). Due to the high similarity between H3 and tH4, the alignment in Figure 3.3 was manually edited to place the insertion after H3 in both TEAD1 and TEAD2/3.

As we were unable to successfully purify any DBD proteins, the implications of the extra region for DNA binding are currently unclear. In the case of human TEAD1 the H3 α-helix is known to be the DNA recognition site (Anbanandam et al. 2006). The structure of the TEA domain is similar to that of the homeodomain fold; DNA binding is also known to be mediated by H3 in homeodomain-containing proteins (Laughon 1991). Given the proximity of tH4 in TEADs 1 and 2/3 to H3, tH4 may affect DNA binding in some way. The presence of an extra helix might enhance or alter the specificity of DNA binding. There are examples of transcription factor DBDs comprising four-helix bundles including E47 (a bHLH TF) (Ellenberger et al. 1994) and Upstream stimulatory factor (USF, a b/HLH/Z TF) (Ferre-D'Amare et al. 1994), however these tend to be dimers of two helix-loop-helix (HLH) structures. Alternatively the extra helix might interfere with or prevent interactions with DNA, or have no effect on DNA binding at all.

The TEAD4 variant with a stop codon before the YBD would presumably be translated as an isolated DNA binding domain. If this assumption is correct, it might be feasible for the non-YAP binding isoform of TEAD4 to competitively inhibit the action of full-length TEAD-YAP thereby preventing transcription. A truncated TEAD would be able to enter the nucleus as the nuclear localisation signal (NLS), "RKQVSSHIQVLARRKAREIQ", identified in the *Drosophila* TEAD homologue Sd (Srivastava et al. 2002) encompasses H3 of the DNA binding domain. The Sd NLS bears between 77% and 94% identity with the medaka TEAD H3 and tH4 helices, so enhanced nuclear localisation might be an alternative role of tH4 in TEAD1 and TEAD2/3.

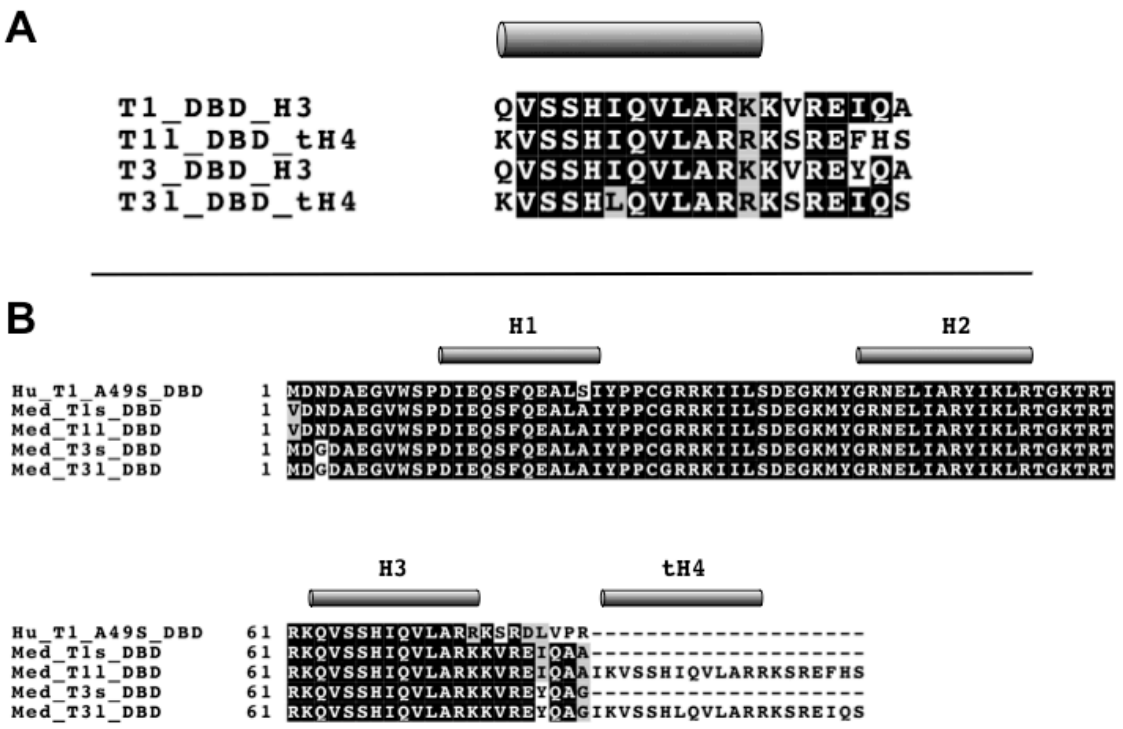


Figure 3.14 (A) Multiple sequence alignment of TEAD1 (T1) and TEAD2/3 (T3) H3 and tH4. The illustration above the alignments highlights the location of residues thought to be located in helical regions based on the structure of human TEAD1 A49S mutant (PDB ID: 2HZD) (Anbanandam et al. 2006). (B) Alignment of the DBD from human TEAD1 A49S with TEAD1 (s and l) and TEAD2/3 (s and l). Alignments were made using ClustalW2 and the panels generated with Boxshade.

3.4.2 Practical aspects of TEAD purification

As TEAD transcription factors are the nuclear target of YAP (and TAZ) regulated by the Hippo pathway, our intention was to purify the DBD of one or more medaka TEAD proteins in order to study TEAD-DNA interactions more closely. At the beginning of the project, the structure of the C-terminal domain of TEAD that interacts with YAP was unsolved. Also, as the amino acid sequence of TEAD (YBD) was not similar to any other protein in the PDB there was no template available for homology modelling. Therefore we intended to study the structure of TEAD and its interaction with YAP.

Unfortunately, it proved extremely difficult to purify sufficient yields of TEAD DBD and YBD protein constructs for biophysical characterisation. Despite using a range of expression methods including changing vectors and cell strains, none of the DBD fragments provided any useful data. Previously, the human TEAD1 (TEF-1) DBD was subjected to site-directed mutagenesis (A49S in the first α-helix based on the *Drosophila* Sd sequence) in order to produce enough protein for structural

characterisation (Anbanandam et al. 2006). This avenue was not explored, though it may be worth considering for potential future experiments.

The best quality ^{1}H-^{15}N-HSQC of the YBD from a TEAD protein was that of TEAD2/3 (239-462) when expressed in the pSV281 vector alongside the pREP4-GroESL chaperonins. The proteins used in this study are derived from medaka fish, a multicellular, vertebrate organism with more complex molecular machinery than the unicellular *E. coli* strains that were used to express the proteins. The co-expression of the bacterial chaperonins GroEL and GroES (on the same plasmid) with proteins from higher organisms was previously shown to improve expression and purification of properly folded constructs for use in NMR experiments (Piserchio et al. 2009). Reproducing a good quality TEAD2/3 NMR sample proved troublesome. Both the TEAD2/3 (YBD) and pREP4-GroESL vectors contain kanamycin resistance genes; therefore it is possible that the cell line lost the pREP4-GroESL plasmid, thereby reducing the ability of the cells to produce well-folded TEAD2/3.

The TEAD4 (YBD) pSV281 construct used for trypsin digestion and NMR experiments was expressed without the assistance of extra chaperones, though as for all successful YBDs, the cells were induced at 12°C and lysed in buffer containing 4% Triton X-100. Although the concentration of TEAD4 (YBD) pSV281 achieved when making the NMR sample was higher than TEAD2/3, the TEAD4 spectrum exhibited signs of heterogeneity such as variable peak intensities. This could be attributed to a misfolded or unfolded fraction of the protein population. Trypsin treatment would have removed unfolded protein but may not have been effective in digesting misfolded protein; production of trypsinised ^{15}N-labelled TEAD4 (YBD) was attempted but acquisition of an improved ^{1}H-^{15}N-HSQC spectrum was not achieved.

Co-expression of TEAD4 (YBD) pOPIN-F with pREP4-GroESL (selection of colonies was made with ampicillin and kanamycin) was used successfully for crystallisation, which suggests the protein was homogeneous. However, this may not have been the case. A fraction of the population could have been misfolded, which might have lead to the crystal twinning effect that will be discussed in section 3.4.3. Again, production of ^{15}N-labelled TEAD4 (YBD) for NMR using this method did not result in a good quality spectrum.

3.4.3 Structural characterisation of TEAD and YAP interaction domains

The trypsin digest of TEAD4 (YBD) gave an initial indication that the domain is globular. The presence of folded secondary structural elements in the YBD of

TEAD2/3 and TEAD4 was confirmed by NMR spectroscopy (Figure 3.6). The ability of the YBD to crystallise was another indication that the protein possesses a rigid secondary structure. NMR data collection was limited due to difficulties in concentrating protein, the poor reproducibility of protein preparation and the size of the domain. This meant that $^{15}N/^{13}C$ samples were not made and triple resonance data sets that might have led to sequence specific peak assignments could not be acquired.

Although TEAD4 crystals diffracted well, at the time there were no homologous models available for molecular replacement to solve the phase problem and calculate an electron density map. Therefore, a host of alternative experimental methods were utilised in an attempt to acquire difference data. Initially the standard method of seleno-methionine labelling was tried in order to obtain phasing data by multi-wavelength anomalous dispersion (MAD). Although protein was produced it would not crystallise, perhaps because of slight differences in sample preparation. The next method attempted was multiple isomorphous replacement phasing (MIR). Various heavy atom salts including mercury (II) chloride, phenyl mercuric acetate and potassium tetrachloroplatinate were soaked into native crystals. Most of the time the additives disrupted the crystal structure to the extent that visible cracks appeared, rendering the crystal useless. The few occasions that the crystal was not visibly damaged, diffraction did not occur or was mosaic, thus indicating sub-microscopic (i.e. not visible under a light microscope) disruption of the crystal lattice.

The TEAD4 (YBD) contains three Cys and six Met residues that could be detected by anomalous sulphur scattering by changing the wavelength of the X-ray beam. Longer wavelength X-rays were used at ESRF (Grenoble) on these crystals, however the data displayed signs of mosaicity and could not be processed, thus the possibility that the crystals were twinned was suggested by one of the course leaders – Christoph Mueller-Dieckmann. In 2010 three independent TEAD (YBD) structures were published: human TEAD2 (Tian et al. 2010), human TEAD1-YAP (Li et al. 2010), and mouse TEAD4-YAP (Chen et al. 2010). As expected, all three structures are highly similar. The key features of the TEAD (YBD) structures are described in Chapter 1. As a native data set had previously been acquired, a final attempt was made to solve the medaka TEAD4 (YBD) structure using molecular replacement with one or more of the published structures as a model for obtaining phases.

Unfortunately it proved impossible to refine the medaka TEAD4 (YBD) structure and the possibility that the crystals might be twinned was reconsidered. This theory was supported by the fact that although indexing the data indicated that the space group was

$P222_1$ (Table 3.4), there was also a high possibility of the unit cell being in the P222 space group. The UCLA merohedral twin detector software indicated that the TEAD4 (YBD) crystals were partially twinned. Twinning is the presence of more than one crystal lattice within what appears to be a single crystal, which leads to overlapping reflections in the diffraction pattern. Twinning may be described as macroscopic, merohedral or non-merohedral. According to CCP4, merohedral twinning gives rise to diffraction patterns that can be superimposed, whereas in the case of non-merohedral twins the reciprocal lattices are not aligned. Macroscopic twins may be seen under the microscope and grow from separate nucleation sites. Not all space groups can form merohedral twins and the two most likely space groups for the TEAD4 (YBD) crystals, P222 and $P222_1$, unfortunately fall into this group. Orthorhombic crystals (like TEAD4) can form 'pseudo-hemihedral' twins (similar to merohedral twins) but only if two of the cell axes are of similar lengths; the criterion for non-merohedral twins also includes two of the cell dimensions being similar. The TEAD4 crystal cell dimensions (Table 3.4) are 63, 90 and 109 Å so they are clearly not pseudo-hemihedral or non-merohedral twins. Therefore, although twinning was not visible at the highest magnification, the TEAD4 crystals must have been macroscopic twins. The CCP4 suite contains the 'DETWIN' programme to correct for merohedral twinning but the software does not currently have the means to refine macroscopic, non-merohedrally or partially twinned datasets. The medaka TEAD4 YBD structure therefore remains undetermined.

3.4.4 TEAD-YAP interaction

A pull-down assay between immobilised His-tagged TEAD4 (YBD) on a nickel ion affinity column and YAP (TBD-WW1) verified that medaka TEAD and YAP interact. Human YAP (TBD) was subsequently shown to bind TEAD2 (YBD) with nanomolar affinity (Tian et al. 2010). Given the many functional studies of the YAP-TEAD interaction, the crystal structures, and the recently discovered co-evolution of YAP and TEAD (Hilman et al. 2011), the results of the pull-down experiment are not unexpected.

Attempts were made to crystallise the YAP-TEAD complex, initially using 96 well high throughput screens and then optimising the conditions in 24 well plates. The scaled-up screens were unsuccessful and after the publication of the 2010 TEAD structure papers (Chen et al. 2010; Li et al. 2010; Tian et al. 2010) the WW domains of YAP and its paralogue, TAZ became the main focus of investigation.

CHAPTER 4: WW DOMAINS OF YAP AND TAZ

(INSIGHTS INTO THE STRUCTURAL, DYNAMIC AND BINDING PROPERTIES OF TANDEM WW DOMAINS FROM THE NUCLEAR EFFECTORS OF THE HIPPO SIGNALLING PATHWAY)

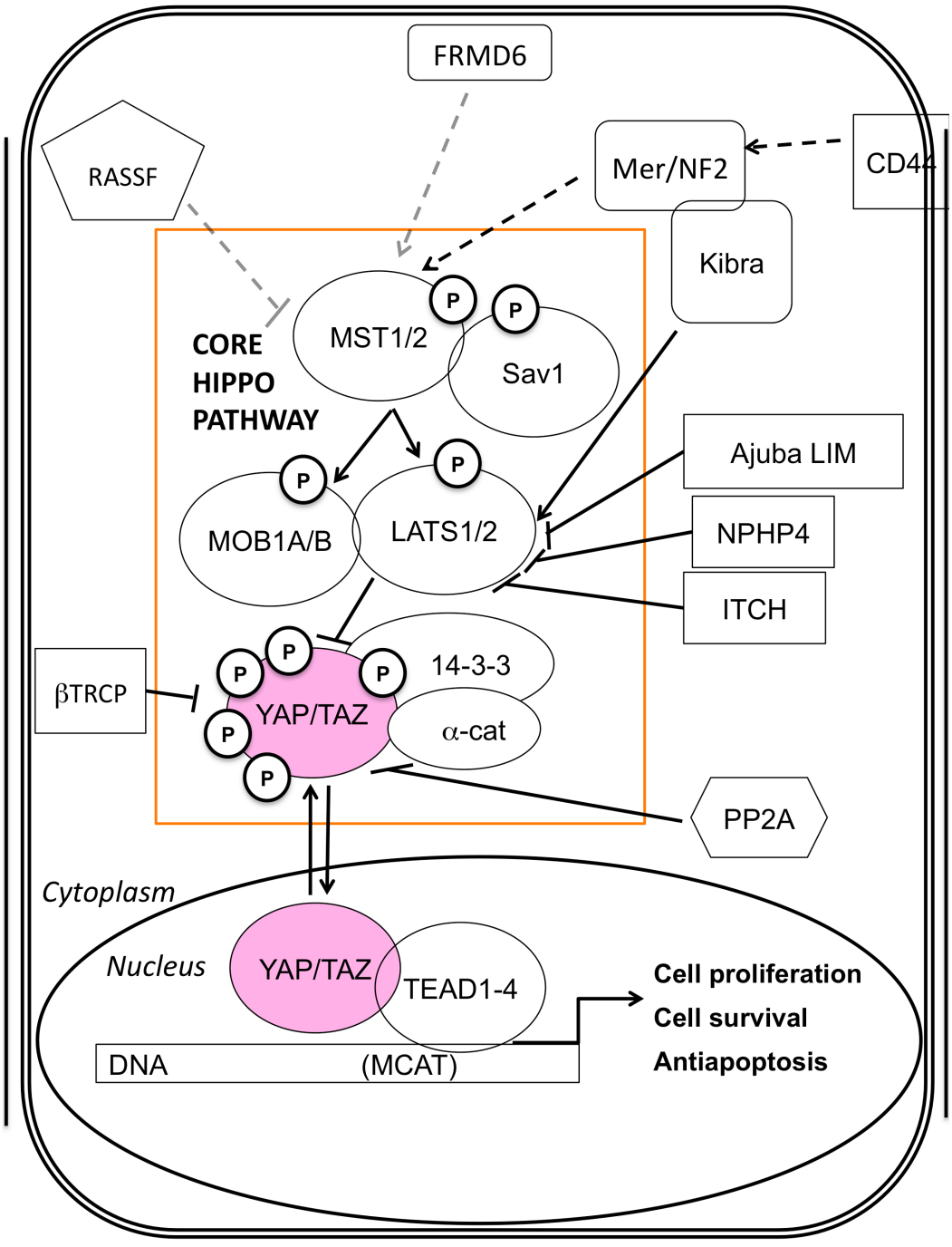


Figure 4.1 Location of YAP and TAZ within the Hippo pathway (highlighted in pink). Core Hippo pathway components are located within the orange box.

4.1 Introduction

WW domains are small protein-protein interaction domains, prevalent in signalling molecules. There are two isoforms of YAP; the presence of either one or two WW domains is the only difference between them. To date only a single WW domain isoform of TAZ has been described. The WW domains of YAP and TAZ are central to their multi-protein binding capabilities (Table 1.2) including the interaction with the LATS kinase that leads to cytoplasmic sequestration of YAP/TAZ by 14-3-3 proteins (Fig 4.1). Previous studies on tandem WW domains have described molecules with a range of different properties including linkers with different lengths and degrees of flexibility, and inter-domain cooperation that leads to significantly enhanced binding affinities. All published tandem WW domain structures to date contain shorter inter-WW linkers (at 12-20 residues) than the inter-WW linker of YAP which is 35 residues. Prior to this work, there were no structure-based studies on the second WW domain or tandem WW1-WW2 domain of YAP.

The solution structure of human YAP WW1 in complex with various peptide ligands has been published (Macias et al. 1996; Pires et al. 2001; Aragon et al. 2011). There are no available solution structures of unliganded YAP WW1 and previous studies have implied that YAP WW1 is unstable in comparison with other WW domains (Ibragimova et al. 1999). Neither were there any structures of YAP WW2. Since the publication of the work in this chapter (Webb et al. 2011), a study of the interaction between a Smad1 peptide and YAP individual and tandem WW domains has been published (Aragon et al. 2011). However, prior to these recent publications, there were no structure-based studies on the WW-pair of YAP or TAZ.

The WW domains of YAP and TAZ are small enough to be characterised by NMR spectroscopy. In addition to gaining insight as to whether a protein is folded (by inspecting the 2D ^{1}H-^{15}N-HSQC spectrum), this technique can be used to generate a model of the 3D structure, study the dynamic properties of a protein and determine which residues are affected by interactions between proteins.

The aims of the work in this chapter were to study the WW domain of medaka TAZ and characterise structural features and peptide-binding properties from the tandem WW domain of medaka YAP. An Ile residue N-terminal to the core WW domain has been implicated in stabilising the structure of WW1 (Macias et al. 1996). In order to determine if this residue played the same role in medaka YAP WW1, both long and short WW1 constructs, WW1 and WW1(s), which contained or lacked the N-terminal

Ile, were created. During the course of this research a second isoform of TAZ (TAZ2) that contains two WW domains was discovered. Structural features and dynamics of the WW domains from medaka YAP and TAZ (Figure 4.2) were characterised by NMR spectroscopy. Functional interactions between the YAP/TAZ tandem WW domain fragments and PPxY motif containing peptides were also revealed by NMR titration experiments.

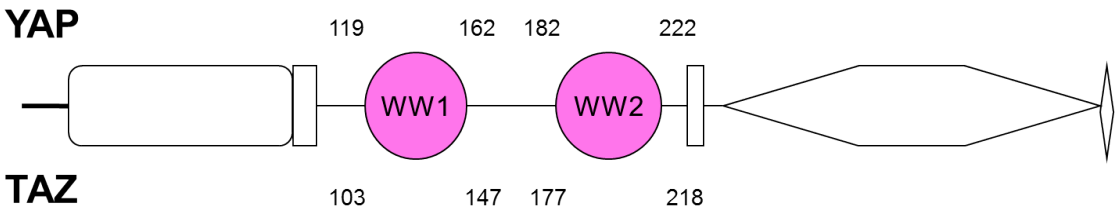


Figure 4.2 Representation of medaka YAP/TAZ domains studied in this chapter. The numbering for YAP and TAZ corresponds to the constructs that were most frequently used. Each individual WW domain is approx. 40 residues in length. The YAP inter-WW linker is approx. 35 residues and the TAZ linker approx. 45 residues. Domains that do not feature in this chapter are unlabelled and do not contain colour.

4.2 Materials and methods

4.2.1 Production of template cDNA and cloning of DNA fragments

Production of TAZ cDNA. The full-length TAZ cDNA plasmids were produced in the same way as TEAD cDNA (section 3.2.1). The first round of PCR was carried out using an annealing temperature of 56°C with the forward and reverse primers 'TAZ 3F' and 'TAZ 3R' (Appendix I-c). The primers for the nested PCR were 'TAZ 4F' and 'TAZ1 innerR' (Appendix I-c), and the annealing temperature was 50°C.

Cloning of WW domains into Vanderbilt vectors. YAP and TAZ WW domain DNA was amplified using Phusion HF DNA polymerase following cloning procedures described in Chapter 2. YAP and TAZ DNA fragments were cloned into the EcoRI and HindIII sites of vectors pSV281, pBG100 and pBG101. Constructs and primers are shown in Table 4.1 and Appendix I.

Table 4.1 YAP and TAZ WW domain cloning information

Protein	Domain	Forward primer	Reverse primer
YAP	WW1 (S119-L162)	CGGAATTCCTCCTACGA GATTCCCGAC	CCCAAGCTTCTATTAGAG CAGGGCTTTGCG
YAP	WW1(s) (D124-L162)	CGGAATTCCGACGACGT GCCCCTG	CCCAAGCTTCTATTAGAG CAGGGCTTTGCG
YAP	WW1-WW2 (S119-R222)	CGGAATTCCTCCTACGA GATTCCCGAC	CCCAAGCTTCTATTAGCG AGTCTCAAGTCG
YAP	WW1(s)-WW2 (D124-R222)	CGGAATTCCGACGACGT GCCCCTG	CCCAAGCTTCTATTAGCG AGTCTCAAGTCG
YAP	WW2 (S182-R222)	CGGAATTCCAGCCCAGC TAGTGGAC	CCCAAGCTTCTATTAGCG AGTCTCAAGTCG
TAZ	WW1 (A103-L147)	CGGAATTCCGCGACATC GATCATACCTGA	CCCAAGCTTCTATTAGAG CTGCGAGAGACGAGG
TAZ	WW1-WW2 (A103-M218)	CGGAATTCCGCGACATC GATCATACCTGA	CCCAAGCTTCTATTACAT TTTCTGAGCTAGACGCG
TAZ	WW2 (N177-M218)	CGGAATTCCAACCCGGA GTCAGGTCCAC	CCCAAGCTTCTATTACAT TTTCTGAGCTAGACGCG

4.2.2 Expression and purification of WW domains from YAP and TAZ

Induction and purification trials were carried out as described in Chapter 2. All YAP and TAZ WW domain constructs were expressed in BL21 (DE3) cells with 0.5 mM IPTG for 3 hours at 37°C, with the exception of YAP WW1 and YAP WW1(s) in pSV281 which were expressed at 12°C overnight in Arctic Express (RP) and BL21 (DE3) cells. ^{15}N and $^{15}N/^{13}C$ labelled preparations were inoculated as described for YAP (TBD-WW1) in Chapter 2. In general, cells were resuspended in His-buffer A with 2% Triton X-100 and one 'Complete mini protease inhibitor cocktail' tablet per 50 ml buffer. Proteins were initially purified by affinity chromatography using a HisTrap FF column (GE Healthcare). The His-tag from pBG100 and the His-GST tag from pBG101 were cleaved overnight at 4°C using 3C protease in 50 mM Tris, 150 mM

NaCl, 1 mM DTT, pH 7.5. Proteins were separated from the cleaved tags by a second affinity purification.

^{15}N and $^{15}N/^{13}C$ labelled YAP WW1-WW2 and TAZ WW1-WW2 were further purified by ion exchange chromatography using a pre-packed QFF Sepharose column (GE Healthcare). The wash buffer for YAP was 10 mM potassium phosphate pH 7, and for TAZ was 20 mM Tris pH 7.5. Proteins were eluted in the wash buffer conditions with a salt gradient up to 0.5 M NaCl. All WW domain constructs were buffer exchanged by concentration/dilution (Table 4.2) and prepared for NMR experiments as previously described (Chapter 2).

Table 4.2 NMR buffer conditions and parameters for YAP and TAZ WW domains

Protein construct	**Buffer**	**Temperature**	**^{15}N SW**	**^{15}N carrier frequency**
YAP WW1 pSV281	50 mM potassium phosphate pH 5.5, 200 mM NaCl, 5 mM EDTA	18 °C	1519 Hz	120.5 ppm
YAP WW1(s) pSV281	50 mM potassium phosphate pH 5.5, 200 mM NaCl, 5 mM EDTA	25 °C	1519 Hz	120.5 ppm
YAP WW1-WW2 pBG101 (cleaved)	50 mM potassium phosphate pH 5.5, 100 mM NaCl, 0.5 mM EDTA	18 °C	1519 Hz	120.5 ppm
YAP WW1(s)-WW2 pSV281	50 mM potassium phosphate pH 5.5, 200 mM NaCl, 5 mM EDTA	25 °C	1519 Hz	120.5 ppm
YAP WW2 pSV281	50 mM potassium phosphate pH 5.5, 200 mM NaCl, 5 mM EDTA	30 °C	1519 Hz	120.5 ppm
TAZ WW1 pBG101 (cleaved)	20 mM Tris pH 7.5, 100 mM NaCl, 5 mM EDTA, 2 mM NaN_3	20 °C	1640.9 Hz	120 ppm
TAZ WW1-WW2 pBG101 (cleaved)	20 mM Tris pH 7.5, 100 mM NaCl	27.5 °C	1640.9 Hz	120 ppm

4.2.3 YAP and TAZ WW domain HSQC backbone assignments

Initially 2D ^{1}H-^{15}N- HSQC spectra were recorded at different temperatures in order to obtain the best quality HSQC. When the optimal temperature was identified an NH_2-only spectrum was recorded identify NH_2 side chain peaks. In order to make sequence specific backbone assignments, 3D HNCO, HNCACO, HNCACB and CBCACONH data sets were recorded on $^{13}C/^{15}N$-labelled proteins. The double and triple resonance data sets were acquired using Varian BioPack pulse sequences. Spectra were processed using NMRPipe/NMRDraw (Delaglio et al. 1995) and analysed with CCPN Analysis (Vranken et al. 2005). ^{1}H, ^{15}N and ^{13}C chemical shifts were referenced to DSS (Wishart et al. 1995).

4.2.4 YAP WW2 solution structure determination

NMR data were acquired and processed as described in section 4.2.3. Additional 3D data sets were acquired and analysed for the assignment of side chain resonances (^{15}N-TOCSY, CCONH, HCCH-TOCSY and HCCH-COSY) and to obtain NOE distance restraints (^{15}N-NOESY HSQC spectra with mixing times of 50 and 200 ms, ^{13}C-NOESY HSQC with a mixing time of 170 ms). Backbone dihedral torsion angles φ and Ψ were predicted using TALOS+ (Shen et al. 2009) within the NMRPipe system. Hydrogen bond distance restraints were predicted from the published YAP WW1 structure (Macias et al. 1996); the NH-O distance was assigned lower and upper values of 1.5 and 2.5 Å, and the N-O distance assigned lower and upper values of 2.5 and 3.5 Å.

Structures were calculated using Xplor-NIH (Schwieters et al. 2006) to perform simulated annealing starting from random extended structures. The VMD-Xplor (Schwieters et al. 2001) visualization package was used to scrutinise ensembles of structures for violated restraints. The final structures were obtained after iterative refinement by removal of consistently violated restraints and using standard Xplor-NIH refinement protocols. Structure quality was assessed using PROCHECK-NMR (Laskowski et al. 1996).

4.2.5 NMR titrations with PPxY peptides

The peptide GTPPPPYTVG was synthesised by Dr Francesca Giuntini (Department of Pharmacy and Pharmacology, University of Bath) and the dual PPxY peptide GTPPPPYTVGGGGGGGTPPPPYTVG was purchased from Thermo Fisher Scientific. Titration data were acquired at the temperatures shown in Table 4.2. In order to monitor WW-peptide binding, ^{1}H-^{15}N-HSQC spectra of YAP and TAZ tandem WW domains were recorded as a function of increasing peptide concentration. The spectra, from each titration point, were overlaid using Sparky (Goddard et al.) or CCPN Analysis (Vranken et al. 2005) to monitor changes in backbone NH chemical shifts. The average chemical shift change of each NH peak was calculated using Equation 4.1. Rieko Ishima (University of Pittsburgh, U. S. A.) calculated the affinities of WW domains for peptides as previously described (Webb et al. 2011).

Equation 4.1: $\Delta\delta_{av}\,(\text{ppm}) = [(\Delta\delta^2\text{HN} + \Delta\delta^2\text{N}/25)/2]^{1/2}$

Where δ is the standard symbol for the chemical shift in parts per million (ppm); δ is calculated using the resonance frequencies (Hz) of the atom of interest and a reference substance (which in this case is DSS), and the frequency of the spectrometer (which in this case is 600 MHz).

A two tailed t-test was carried out to determine whether peptide binding affected one of the WW domains within the WW1-WW2 fragment significantly more than the other. In order to assess the potential multimerisation effect of dual PPxY peptide binding to the tandem WW domain fragments, native-PAGE was carried out in a similar manner to SDS-PAGE (described in Chapter 2), but SDS and β-mercaptoethanol were omitted from all buffers, the current was reduced to 40mA and the tank was placed on ice to avoid heat denaturation of the protein.

4.2.6 NMR relaxation and dynamics

^{15}N YAP and TAZ WW1-WW2 domains were buffer exchanged into water by dialysis and freeze dried prior to shipping. Rieko Ishima (University of Pittsburgh, U. S. A.) acquired and analysed the ^{15}N NOE, R_1 and R_2 and relaxation dispersion data on the unbound tandem WW domain protein fragments as described in Appendix IV on pages

3301 and 3302 of the article (Webb et al. 2011). Equivalent data for dual PPxY peptide-titrated WW1-WW2 fragments were collected on the in-house Varian INOVA 600 spectrometer.

4.3 Results

4.3.1 TAZ cDNA isoforms

Several full-length TAZ cDNA plasmids were sequenced (Appendix II-e) and two splice forms were identified. The single WW domain isoform, now called TAZ1, was previously known to exist. The identification of a tandem WW domain isoform, TAZ2 is novel, but perhaps not surprising given the existence of the two YAP splice forms (YAP1 and YAP2). Comparison of medaka YAP2 and TAZ2 WW domains reveals a difference in the length of the inter-domain linker of 10 residues (35 residues for YAP2 and 45 residues for TAZ2). The tandem WW fragments shown in Figure 4.3 are 59% identical and 70% similar. The individual WW1 domains of YAP and TAZ have 75% identity and 86% similarity. The individual WW2 domains of YAP and TAZ are 75% identical and 90% similar.

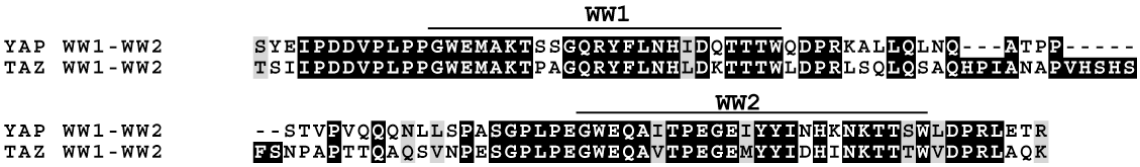


Figure 4.3. YAP2 and TAZ2 tandem WW domain alignment made using ClustalW (Thompson et al. 1994). The panel was generated in Boxshade. YAP residues start at S119 and end at R222. TAZ residues start at T104 and end at K217.

4.3.2 Expression and purification of YAP and TAZ WW domains

YAP WW1 was induced overnight at 12°C in Arctic Express (RIL) cells, as expression was unsuccessful in BL21 (DE3) cells at 37°C. All other YAP and TAZ WW domains were expressed in BL21 (DE3) cells at 37°C for 3 hours. Individual YAP WW domains were generally expressed in the pSV281 vector and purified by affinity chromatography using the N-terminal 6His-tag, which was not usually removed. YAP WW1-WW2 was purified as illustrated in Figure 4.4 and TAZ was purified in the same way. Individual TAZ WW domains were expressed in the pBG101 vector. The His-

GST tag was cleaved and the TAZ proteins purified with similar elution profiles to those shown in Figure 4.4a and 4.4b.

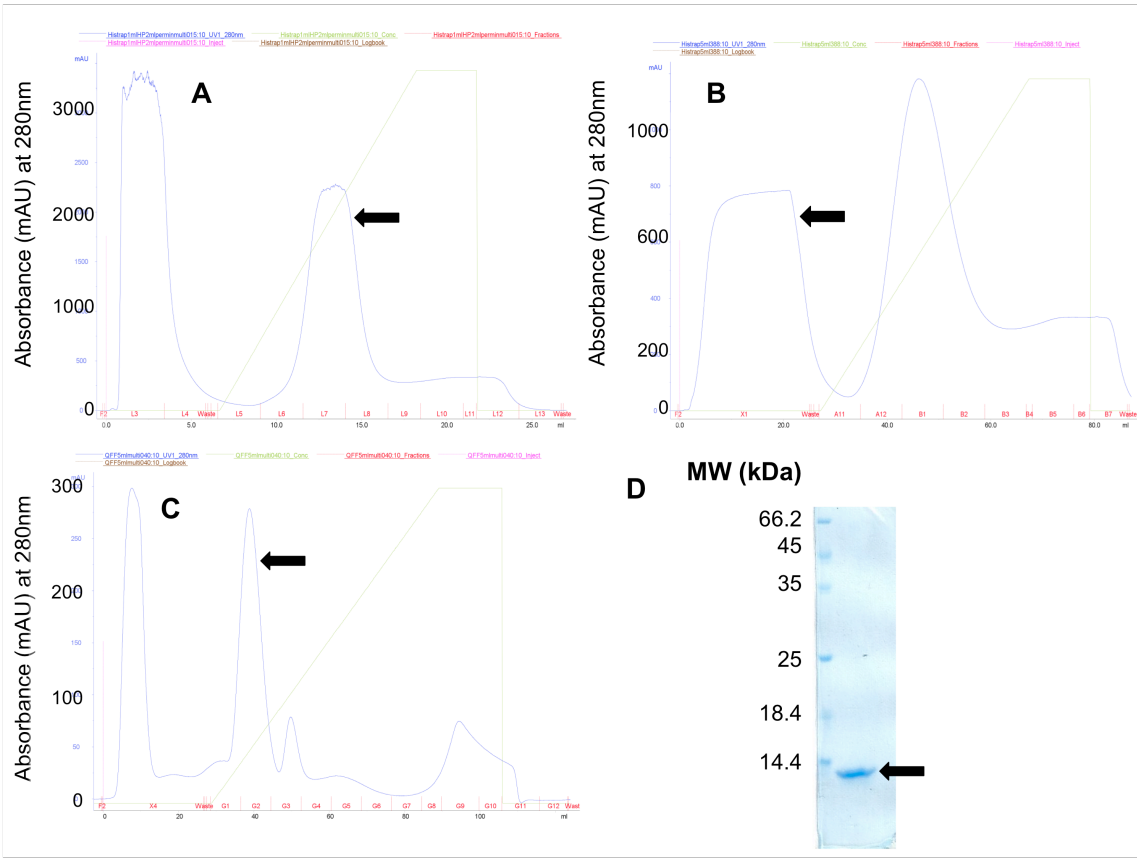


Figure 4.4 Purification profiles of ^{15}N YAP WW1-WW2 from a 0.75 L culture used in NMR experiments. Nickel ion affinity chromatograms of (A) pre- (10 runs using a 1 ml His-Trap FF column) and (B) post- (3 runs using q 5 ml His-Trap FF column) His-GST tag cleaved protein. (C) Ion exchange chromatogram (4 runs using a 5 ml QFF column in KH_2PO_4 pH7 buffer and eluted in the same buffer with the addition of up to 0.5 M NaCl) of post-cleaved, post-affinity purified sample. (D) Coomassie stained SDS-PAGE gel containing the 1 μl of the final NMR sample, with an expected size of ~12 kDa. The pure YAP WW1-WW2 fraction is highlighted in each frame by an arrow.

4.3.3 YAP WW1(s) vs. YAP WW1

It was previously reported that an Ile residue N-terminal to the human YAP WW1 domain was essential to folding (Macias et al. 1996). We also observed this when comparing medaka WW1(s) and WW1 HSQC spectra (Figure 4.5a). The YAP WW1(s) construct starts 4 amino acids after this key Ile and the spectrum shows many peaks around 7.5-8.5 ppm in the ^{1}H dimension but none further upfield or downfield. In

contrast, the WW1 construct that starts two residues N-terminal to I121 produces a ^{1}H-^{15}N-HSQC with well dispersed peaks, suggestive of a folded domain.

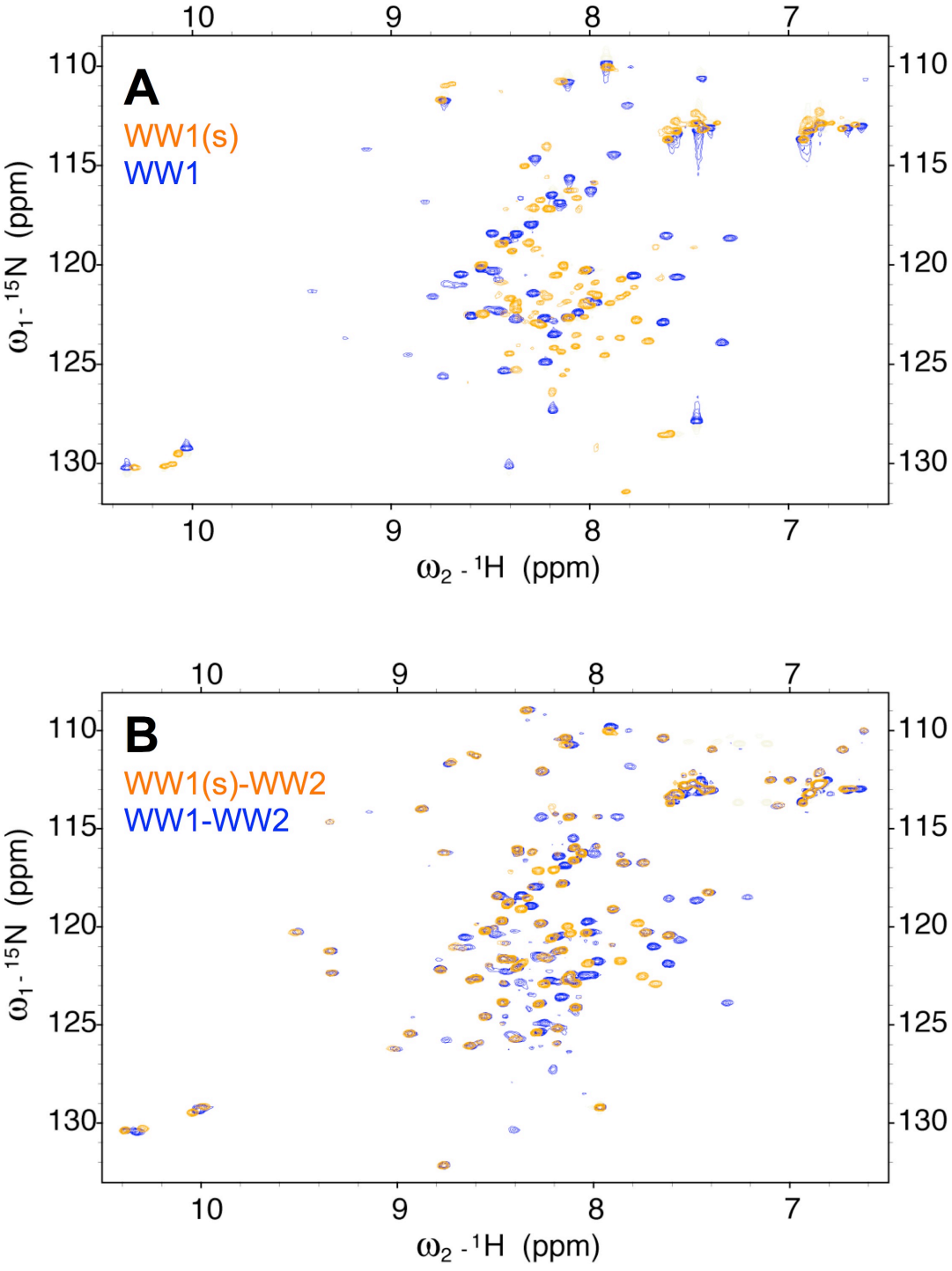


Figure 4.5 ^{1}H-^{15}N-HSQC spectra of (A) WW1 (blue) and (B) WW1-WW2 (blue), which both contain I121, overlaid onto (A) WW1(s) (orange) and (B) WW1(s)-WW2 (orange), which both lack I121, spectra to emphasise the effect of I121 on the structure of YAP WW1 (A) alone and (B) within WW1-WW2. Data were acquired at 25°C at pH 5.5 (buffer conditions in Table 4.2 in section 4.2.2).

4.3.4 Assignment of YAP WW1, WW1-WW2 and WW2 NMR spectra

Making assignments. In order to make sequence specific backbone assignments, a range of 3D NMR spectra were recorded on double labelled protein (^{15}N and ^{13}C) including (Figure 4.6): HNCACB which detects the chemical shift of the Cα and Cβ; CBCACONH which detects the Cα and Cβ of the preceding residue in the sequence; HNCACO which detects the C=O chemical shift; and HNCO which detects the C=O of the preceding residue in the sequence. The ^{13}C backbone C=O chemical shifts are in a different region to those of the side chain carbon atoms. Typically the C=O chemical shifts are around 166-182 ppm, Cα and Cβ chemical shifts are 15-75 ppm; The atoms of each amino acid have characteristic chemical shifts within these ranges (which can be found on the website of the Biological Magnetic Resonance Data Bank-BMRB) and so can often be identified.

WW domain NMR spectra assignments. WW1-WW2 backbone NH peak assignments (Figure 4.7) required the individual WW1 chemical shift assignments (Figure 4.8) for completion, due to weak/undetectable WW1 (within WW1-WW2) cross-peaks in 3D spectra. NMR experiments on constructs containing WW1 (Figures 4.7, 4.8) were recorded at a lower temperature than those of individual WW2 (Figure 4.9) because there were fewer WW peaks at temperatures higher than 18°C, whereas the best quality WW2 ^{1}H-^{15}N HSQC spectra were obtained at 30°C. WW1 peaks around 9-10 ppm, in the 2D ^{1}H-^{15}N-HSQC of WW1 and WW1-WW2, are less intense than the other peaks in the spectrum indicating some heterogeneity within the population. The equivalent WW2 peaks around 9-10 ppm are more intense than the WW1 peaks, and are of similar intensity to the rest of the spectrum, thus suggesting that WW2 adopts a stable folded conformation. Individual WW1 and WW2 chemical shifts are very similar to those within WW1-WW2 (Figures 4.7-4.9), thus implying that the two domains do not interact.

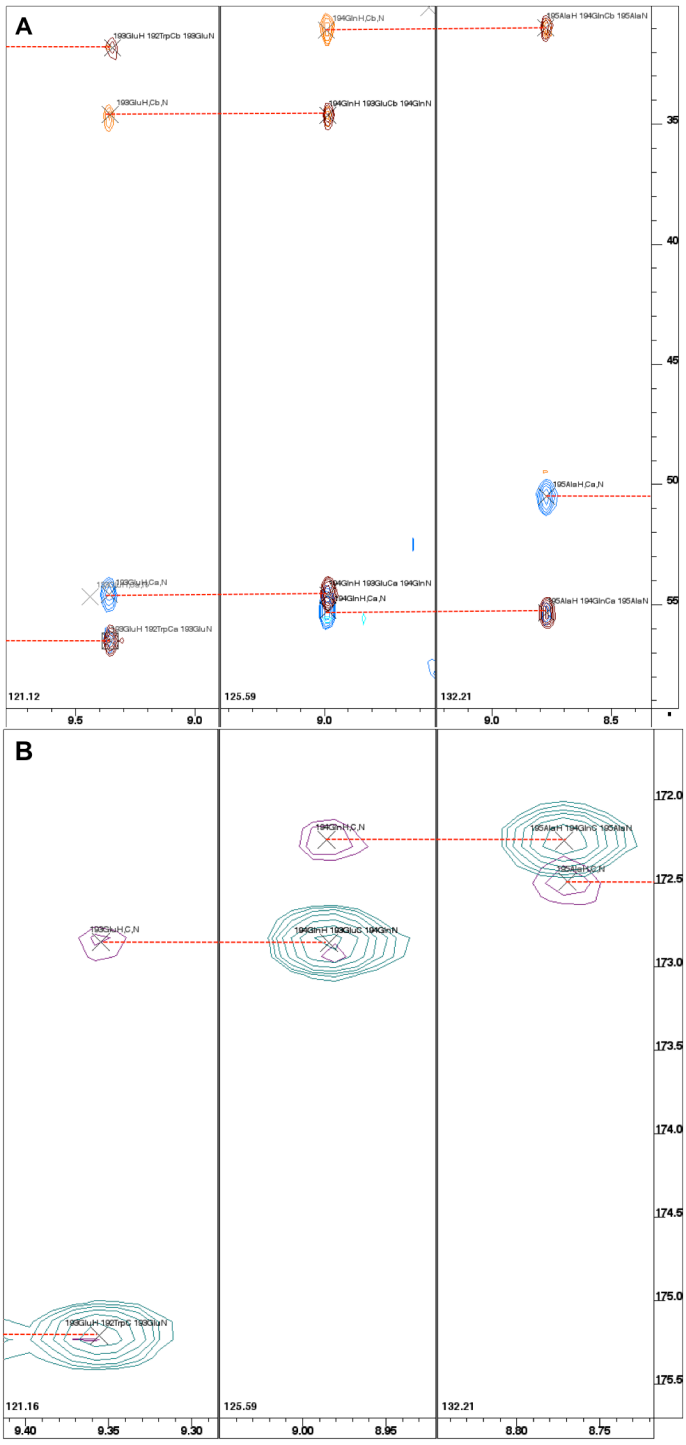


Figure 4.6 Example of cross peak connections for sequence-specific chemical shift assignment. (A) HNCACB and (B) HNCACO show Cα (blue), Cβ (orange) and CO (purple) chemical shifts of the residue of interest (Q194 in the central strip, E193 to the left and A195 to the right). (A) CBCACONH and (B) HNCO show Cα (maroon), Cβ (maroon) and CO (green) chemical shifts of the preceding residue in the sequence. Smaller signals corresponding to the chemical shift of the preceding residue in the sequence are also visible in the HNCACB and HNCACO spectra. ^{13}C chemical shift values are in the y-axes, ^{1}H chemical shifts are in the x-axes, and ^{15}N chemical shifts are in the bottom left corner of each strip. Dotted red lines illustrate the connectivity between residues.

Initial analysis of the assigned YAP WW1-WW2 HSQC indicates that the WW domains contain structural elements and the linker region is natively unfolded. The linker residue peaks are not located in well dispersed areas of the HSQC spectrum, they are clustered around 8-8.5 ppm in the ^{1}H dimension. Conversely, many WW1 and WW2 peaks are dispersed beyond this central region of the HSQC spectrum.

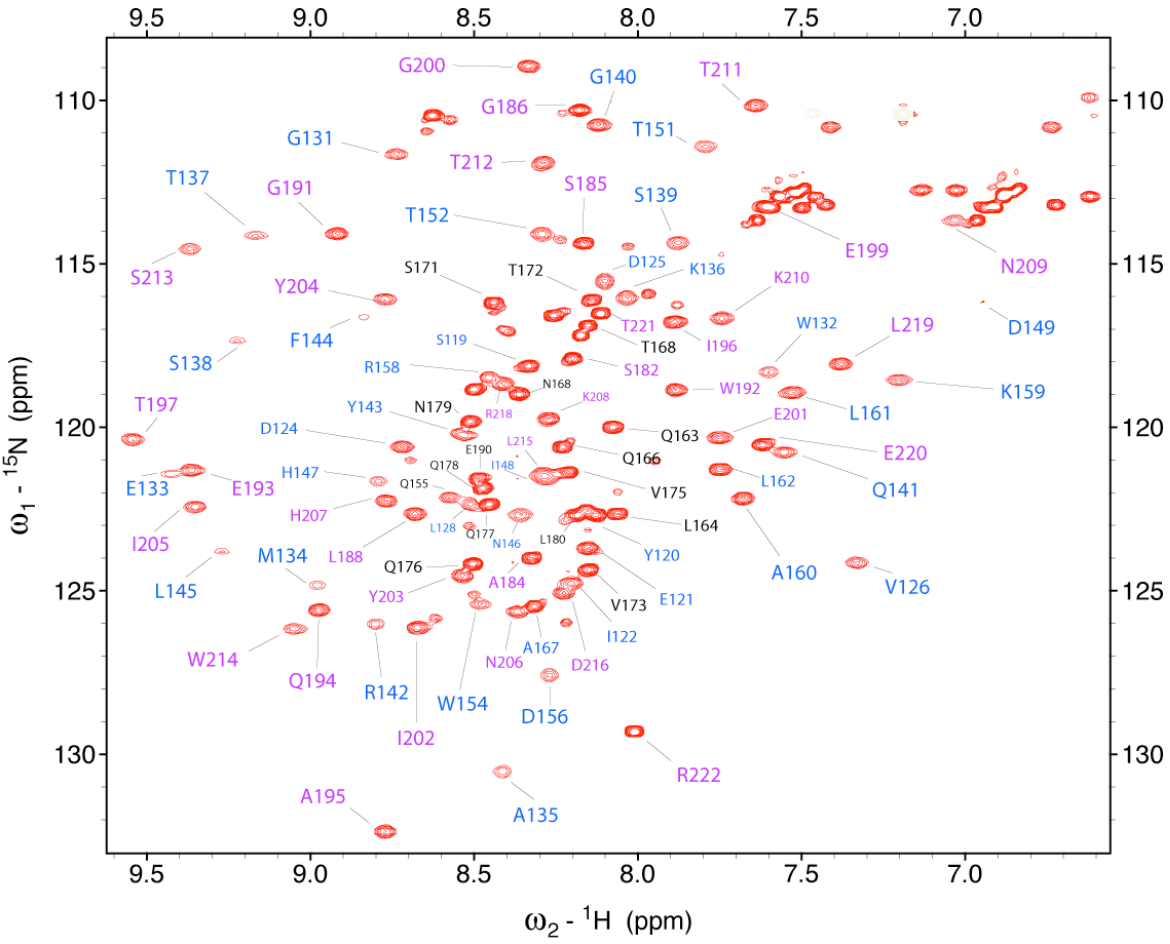


Figure 4.7 YAP WW1-WW2 ^{1}H-^{15}N-HSQC at 18°C with NH backbone peak assignments labelled. In order to facilitate label visualisation, Trp indole NH peaks at 10-11 ppm are not shown. Chemical shifts from WW1 are labelled in blue, linker chemical shifts have black labels, and WW2 peaks have purple labels. All linker residues are located in the centre of the spectrum indicating that the linker is intrinsically unstructured.

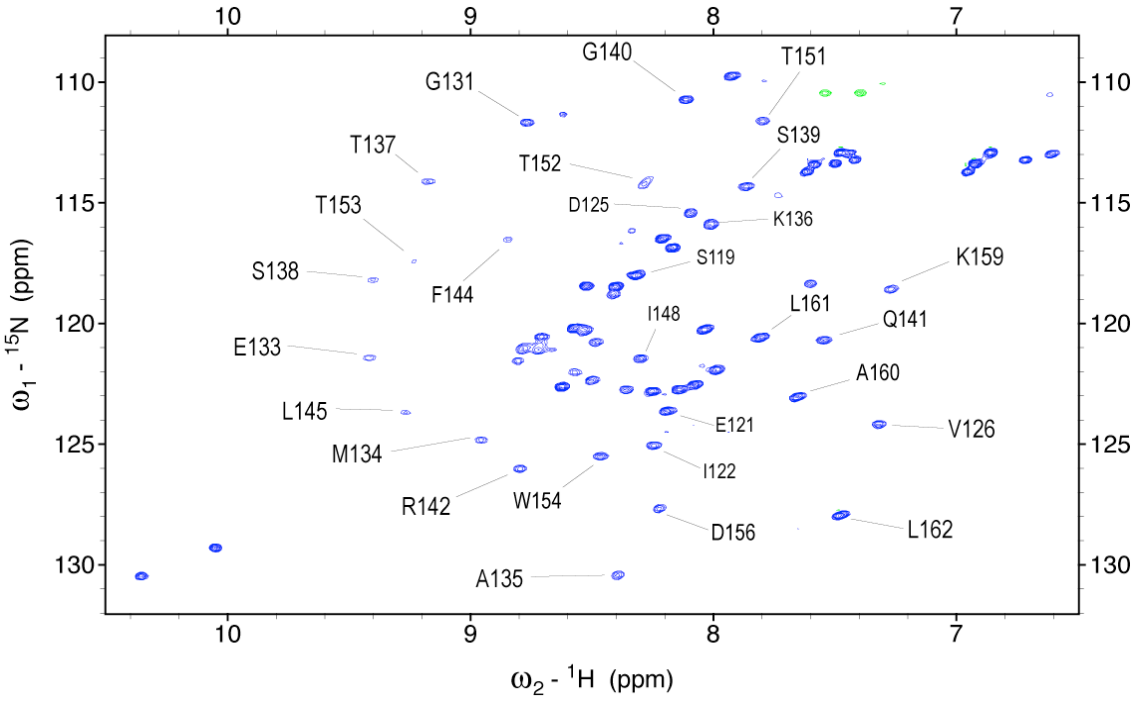


Figure 4.8 His-tagged YAP WW1 ^{1}H-^{15}N-HSQC at 18°C with labelled NH backbone peak assignments.

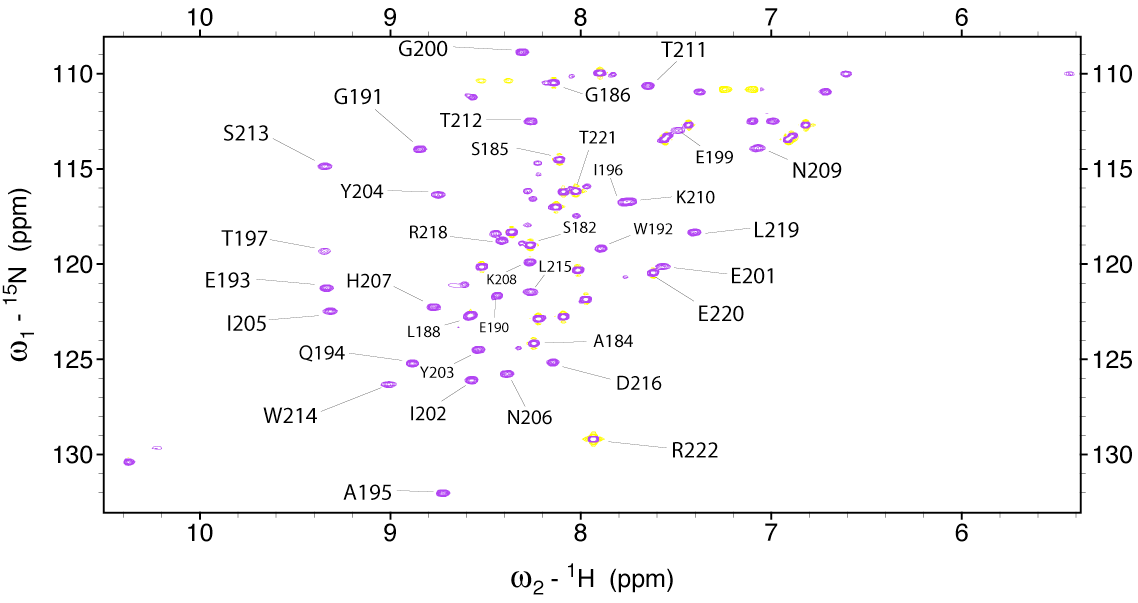


Figure 4.9 His-tagged YAP WW2 ^{1}H-^{15}N-HSQC at 30°C with NH backbone peak assignments labelled.

4.3.5 Solution structure of YAP WW2

The 3D solution structure of WW2 was determined using NMR data (as described in section 4.2) in order to confirm the presence of a typical WW domain fold. That WW2 proved to be more stable than WW1 and WW1-WW2 enabled completion of the task. WW1 alone and within WW1-WW2 was not stable enough to allow the structures of these constructs to be solved. The calculated structure ensemble (Figure 4.10, Table 4.3) verifies that the domain consists of a triple-stranded, twisted β-sheet with two short flexible loops, which is characteristic of WW domains. The root mean squared deviation (r.m.s.d.) in Table 4.3 shows the average distance between the atoms of the 20 superimposed structures. As expected, the flexible regions N- and C-terminal to the β-sheet arch around the hydrophobic face of the WW domain in a 'buckle'-like manner. Opposite the hydrophobic face lies the anticipated peptide-binding site, the activity of which is reportedly dependent upon the presence of the Trp residue in the third β-strand (Koepf et al. 1999).

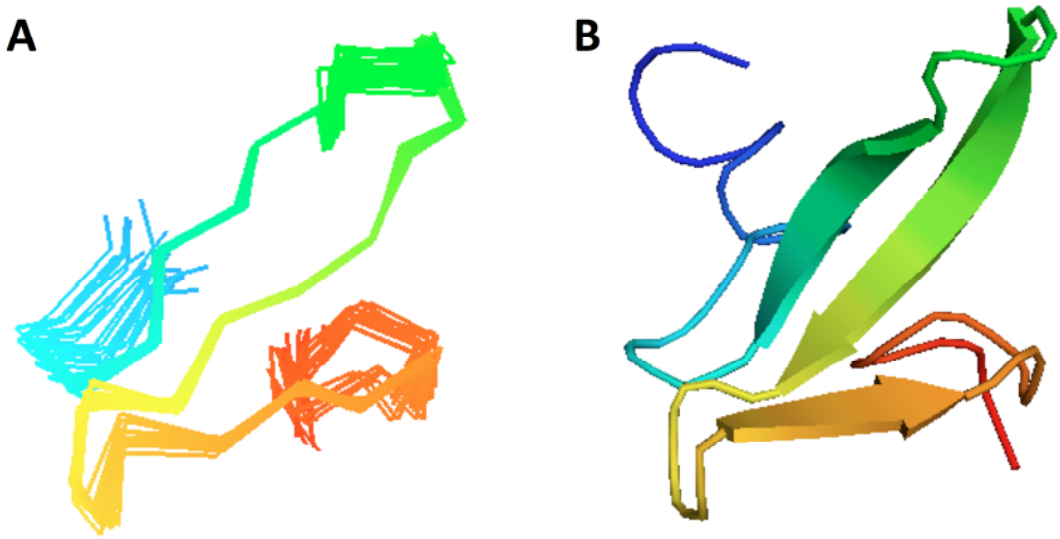


Figure 4.10 Solution structure of medaka YAP WW2 (PDB ID: 2L4J). (A) Superimposition and (B) average structure of the 20 lowest energy structures coloured from blue at the N-terminus to orange at the C-terminus. The figure was taken from Webb et al (2011). The core WW domain contains the canonical triple-stranded β-sheet fold; the N-terminus, which corresponds to the end of the inter-WW linker, in (B) appears to be almost helical, however the orientation of this coiled region is not superimposed within the ensemble. The beginning of the WW linker (C-terminal to WW1) is not shown in any WW1-peptide structures within the PDB and the 35 residue YAP linker region is also not shown in any structures within the PDB.

Table 4.3. Structural statistics summary for the 20 lowest energy structures of YAP WW2. Adapted from Webb et al (2011).

Total number of NOE restraints	358
Intra-residue	108
Sequential and medium range (i+1 to i+4)	113
Long range	137
Number of dihedral angle restraints	46
Number of hydrogen bond restraints	20
Mean r.m.s.d. of backbone atoms*	0.34 Å
Mean r.m.s.d. of non-hydrogen atoms*	1.04 Å
Average number of NOE violations	1.8
Average number of dihedral angle violations	0
Ramachandran plot regions (% residues)+	
Most favoured	81.8
Additionally allowed	18.2
Generously allowed	0
Disallowed	0

* The r.m.s.d. from the mean structure calculated over residues 190-197, 201-207 and 210-214.

+ Calculated using PROCHECK-NMR (Laskowski et al. 1996) over residues 190-214

4.3.6 Assignment of TAZ WW domain NMR spectra

TAZ WW domain ^{1}H-^{15}N-HSQC spectra display more heterogeneous signal patterns than YAP spectra, such as broader line widths and at least one extra Trp side chain NH peak around 10 ppm that indicates a certain degree of chemical exchange. Like YAP, TAZ also required the assignment of WW1 peaks (Figure 4.11) in order to obtain the full complement of WW1-WW2 backbone NH peaks (Figure 4.12). Another similarity to the YAP tandem WW domain spectrum is the apparent lack of structure in the linker region. NMR data were not acquired for TAZ WW2.

^{1}H, ^{15}N and ^{13}C chemical shift based structure prediction was carried out for TAZ WW1 and WW1-WW2 using TALOS+ (Shen et al. 2009). WW1 alone and WW2 within WW1-WW2 retain the three β–strands of the WW domain fold as described for the YAP WW2 structure. Due to incomplete Cα and Cβ assignments for WW1 within WW1-WW2 only the presence of the second β-strand was confirmed. However, NH chemical shifts are conserved between the individual and tandem WW1 HSQC spectra and TAZ WW1 has high sequence homology to YAP WW1, therefore TAZ WW1 within the tandem domain is also likely to fold in a similar manner.

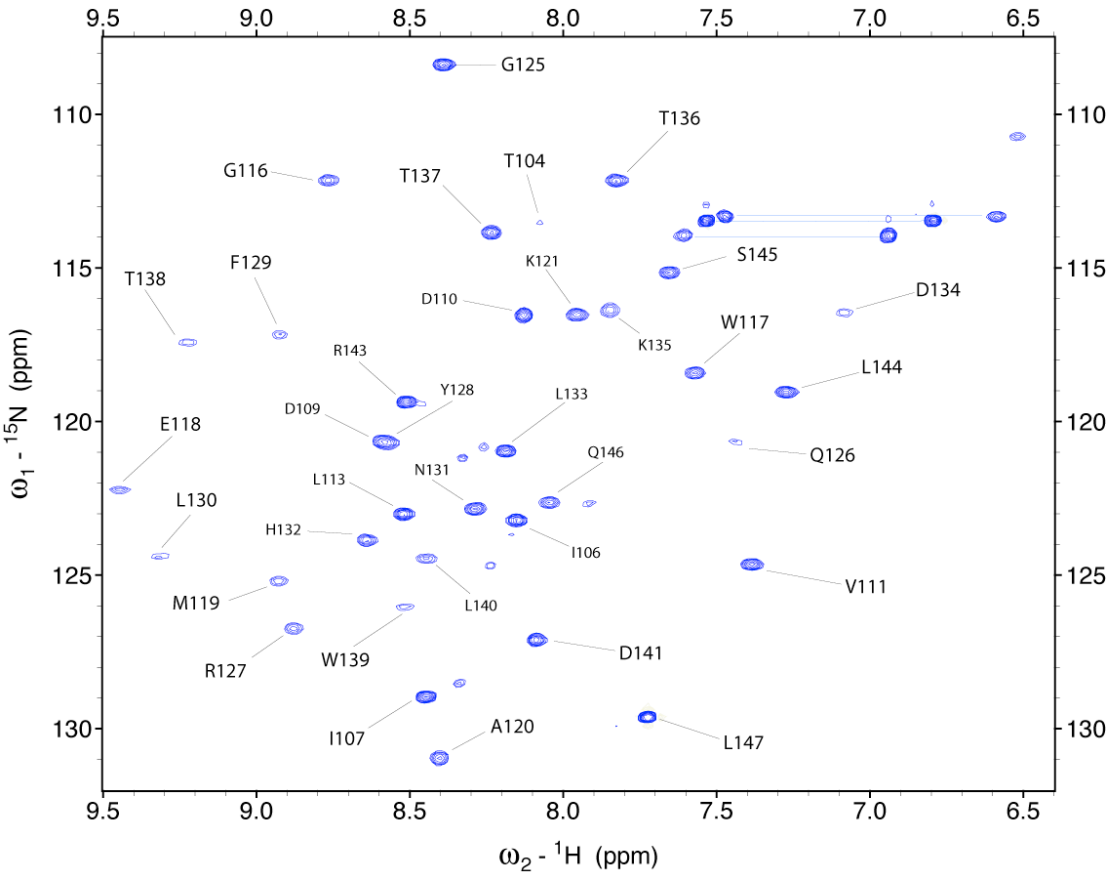


Figure 4.11 TAZ WW1 (103-147) ^{1}H-^{15}N-HSQC spectrum with backbone NH peak assignments. Data were acquired at 20°C. Trp indole peaks around 10 ppm have been omitted to facilitate labelling.

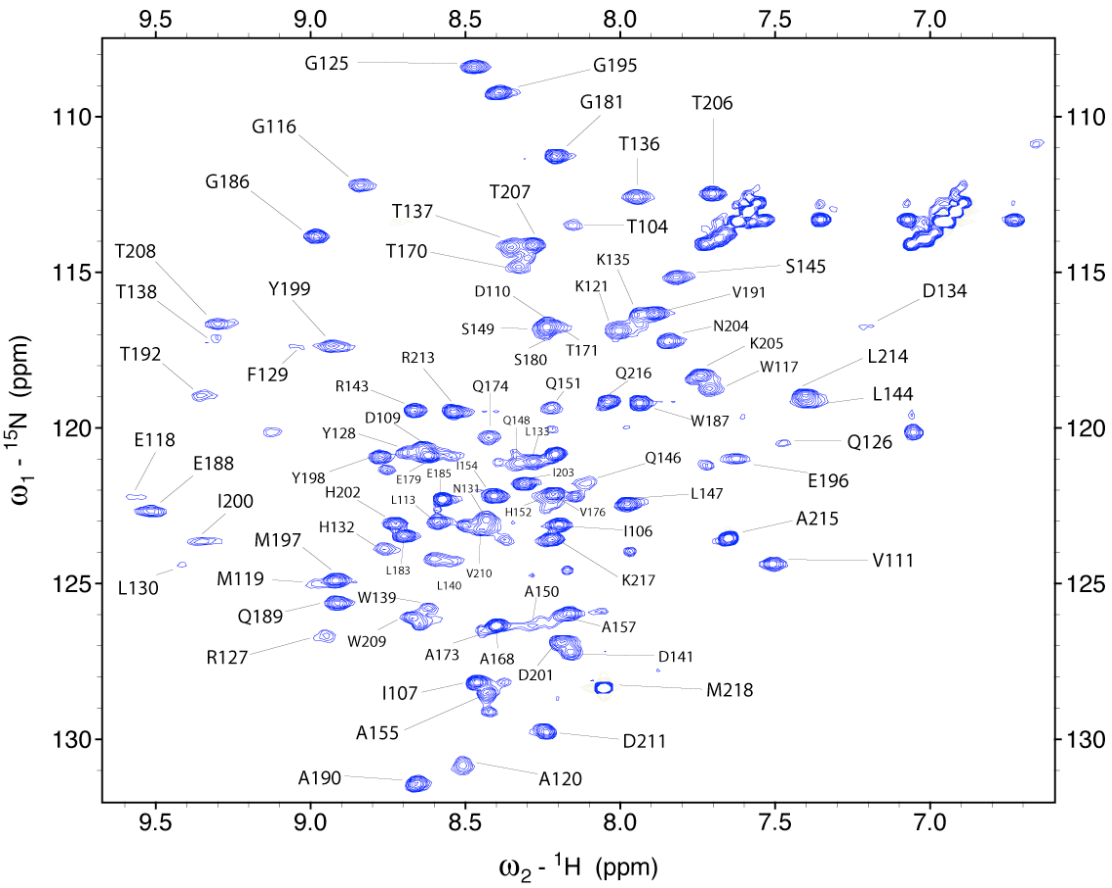


Figure 4.12 TAZ WW1-WW2 (103-218) ^{1}H-^{15}N-HSQC spectrum with backbone NH peak assignments. Data were acquired at 27.5°C. Trp indole peaks around 10 ppm are not shown.

4.3.7 WW1-WW2 NMR titrations with single and dual PPxY peptides

Both YAP and TAZ tandem WW domains were titrated with the single PPxY containing peptide GTPPPPYTVG that has been used in previous interaction studies with human YAP WW1 (Macias et al. 1996), and a tandem repeat version GTPPPPYTVGGGGGGGTPPPPYTVG. Both WW domains of both proteins have the ability to bind both peptides as evidenced by the movement of WW domain peaks in the ^{1}H-^{15}N-HSQC spectra recorded at each titration point (Figures 4.13 and 4.14). Peaks that underwent large chemical shift changes often broadened to the point where identification was difficult. Also, a number of peaks split into two during the dual PPxY titration, indicating slow chemical exchange. Linker peaks underwent little to no movement, suggesting that this region does not interact with, and remains structurally unaffected by, the presence of either peptide.

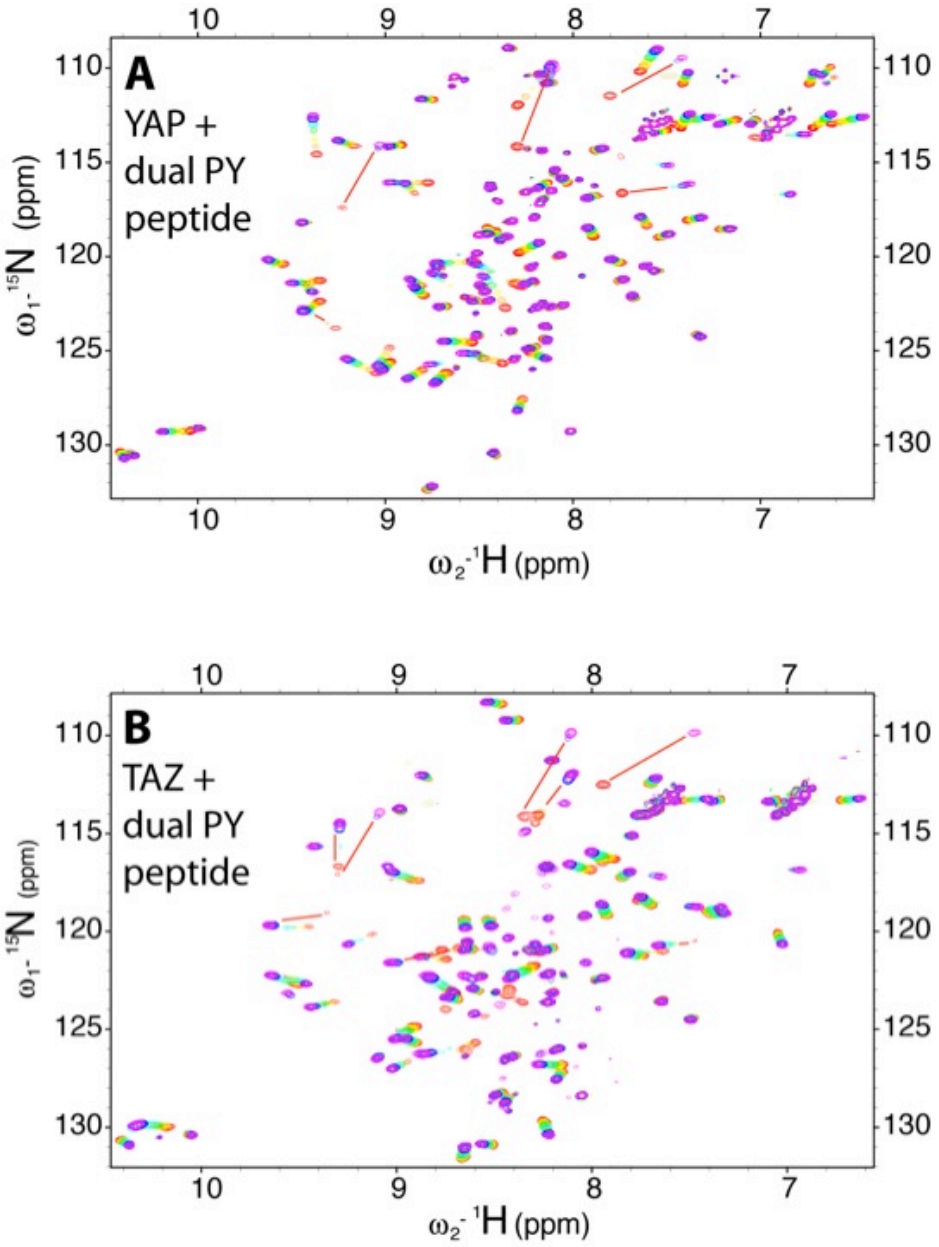


Figure 4.13 Titration of (A) YAP WW1-WW2 and (B) TAZ WW1-WW2 with dual PPxY peptide. Overlaid ^{1}H-^{15}N-HSQC spectra illustrate peak movements of the tandem WW domains without peptide (red) through to the final addition of peptide (magenta). A red line illustrates peak movement in the cases where intermediate titration point peaks have broadened to invisibility. NMR spectra with single PPxY peptide were similar to the dual PPxY titration spectra, but fewer peaks disappeared during the course of the titration.

WW1 peaks exhibit larger chemical shift changes upon binding to peptide than those of WW2 (Figure 4.14). The intensity of WW1 peaks in the 9-10 ppm proton chemical shift region of the ^{1}H-^{15}N-HSQC spectra also increase upon binding to peptide. A paired two-tailed t-test was performed on the overall chemical shift changes in the individual WW domains within the tandem domain constructs. A significant difference is revealed

in overall peak movement between WW1 and WW2 for both YAP and TAZ WW1-WW2 when binding to each peptide. To determine whether the difference in chemical shift movements were related to affinities of the WW domains for peptide, K_d values were calculated using the average chemical shift movement at each titration point for residues within the known binding site.

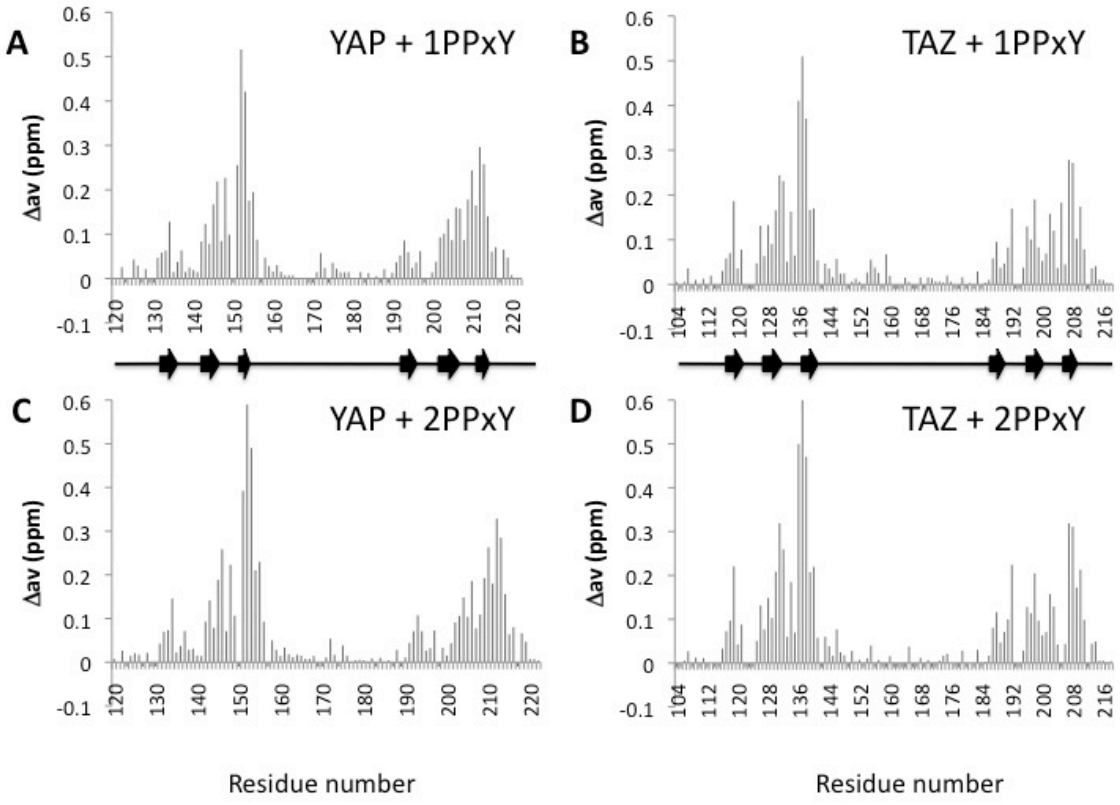


Figure 4.14 Average chemical shift movement of each assigned peak after the final addition of single PPxY peptide (top) and dual PPxY peptide (bottom) to (A, C) YAP and (B, D) TAZ WW1-WW2. Figure taken from Webb et al (2011).

The two WW domains within WW1-WW2 have similar affinities for single PPxY peptide in YAP and TAZ, with TAZ having a slightly lower affinity for the peptide than YAP. YAP has WW1 and WW2 K_d values of 284 +/- 52 μM and 281 +/- 55 μM, and TAZ has WW1 and WW2 K_d values of 330 +/- 116 μM and 273 +/- 116 μM. The similarity between WW1 and WW2 K_d values indicates that the more significant change in the chemical environment of WW1 is not due to a difference in affinity, but a larger effect on the overall structure of the WW1 domain.

Titration of WW1-WW2 constructs with the dual PPxY peptide resulted in chemical shift changes at much lower peptide concentrations than the titration with single PPxY

peptide. For TAZ the K_d was calculated to be at least 6 times higher (48.5 +/- 12 μM) than that for the single PPxY peptide, and the YAP K_d could not be accurately calculated with confidence but is estimated to be similar to that of TAZ. The pattern of WW domain chemical shift changes was similar for single and dual PPxY peptides (Figure 4.14 and 4.15), suggesting identical binding mechanisms with both ligands.

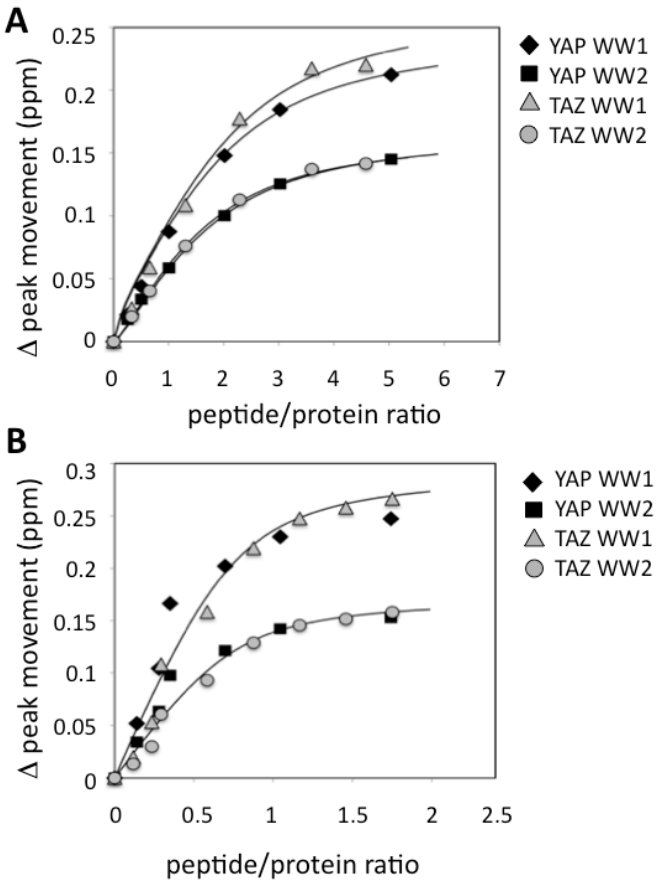


Figure 4.15 Average chemical shift movement of binding site residues from YAP/TAZ tandem WW domains upon interaction with: (A) single and (B) dual PPxY peptide. Figure taken from Webb et al (2011).

If the equipment and an excess of peptide had been available, the affinities of YAP and TAZ tandem WW domains for the peptides would have been measured by isothermal titration calorimetry to confirm the NMR calculations. Attempts were made to monitor the WW-peptide interactions through fluorescence spectroscopy; both by monitoring Trp fluorescence which would, in theory, change upon the interaction between the second W and the PPxY, and also using 1-anilinonaphthalene-8-sulfonic acid (ANS). Unfortunately the fluorescence experiments were unsuccessful as very little difference was observed between WW with peptide and without.

4.3.8 NMR relaxation and dynamics of tandem WW domains from YAP and TAZ

Most of the relaxation and dynamics data were acquired and analysed by Rieko Ishima (University of Pittsburgh, U. S. A.). The exception being relaxation data acquired in-house on the post-titration dual PPxY-YAP WW1-WW2 sample. The results of the unbound tandem WW domain (Figure 4.16) and YAP WW1-WW2-dual PPxY relaxation (Figure 4.17) experiments are included here to present the whole picture as described in Webb et al (2011) (Appendix IV).

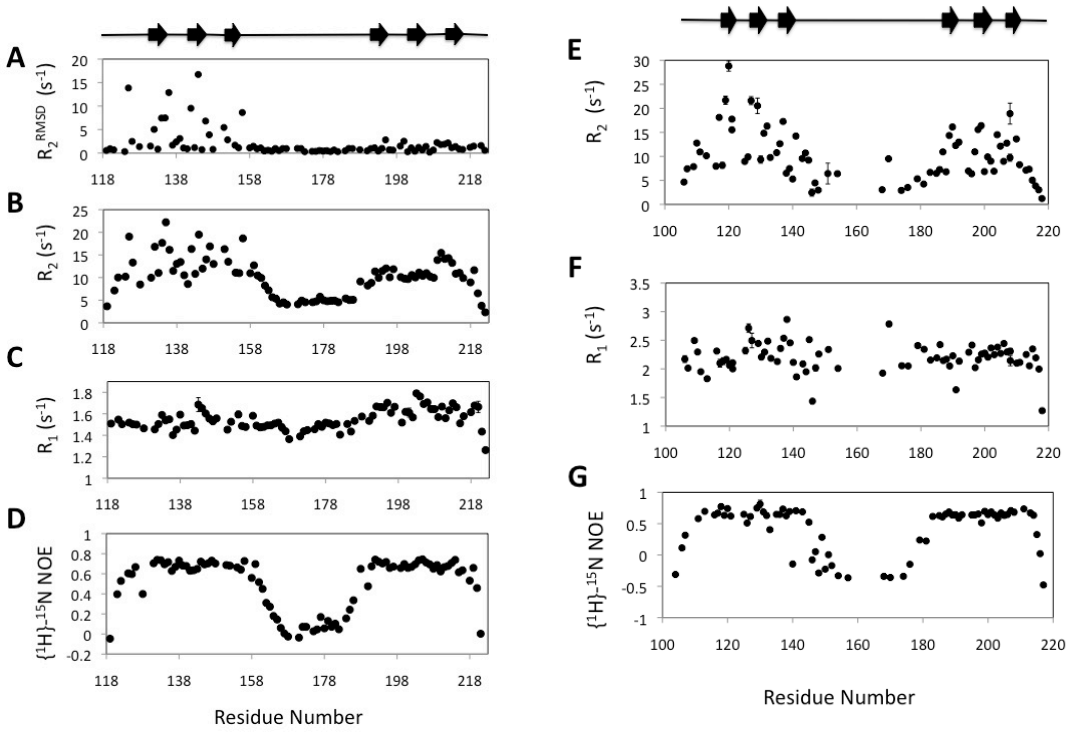


Figure 4.16 (A-D) YAP and (E-G) TAZ tandem WW domain backbone dynamics and relaxation data, adapted from Webb et al (2011). The β-strands shown for YAP are from published WW domain structures (Macias et al. 1996; Webb et al. 2011) and those shown for TAZ were predicted in TALOS+ using chemical shift values.

The relaxation rates R_1 and R_2 are calculated from the relaxation times T_1 and T_2, which measure the time taken (μs-ms) for an excited atom to decay back to thermal equilibrium (T_1) and to lose spin coherence (T_2). Large and fluctuating R_2 values signify conformational and/or solvent exchange in a dynamic region of a protein. Additional dynamic information can be elucidated on the ps-ns timescale using {^{1}H}-^{15}N-NOE data, which indicate the degree of movement in the backbone NH bond. High values indicate restricted movement characteristic of a rigid folded region and low values

suggest a higher degree of flexibility. Relaxation and dynamics experiments required ^{15}N labelled protein and detected the chemical shifts of the backbone amides.

The YAP R_2 dispersion i.e. the transverse relaxation rate, R_2, measured as a function of the effective field using a Carr-Purcell-Meiboom-Gill (CPMG) pulse train (Figure 4.15), indicates that YAP WW1 undergoes significant conformational exchange. The side chains of the residues involved in chemical exchange, as determined by R_2 RMSD, lie on the opposite face of the β-sheet to the peptide-binding site and comprise D125, W132, M134, A135, K136, R142, F144, T151 and D156. This correlates with the weak WW1 ^{1}H-^{15}N-HSQC peak intensities observed in Figures 4.7-4.8. Both WW domains of TAZ display signs of conformational instability with high CPMG R_2 values, consistent with the signs of heterogeneity observed in ^{1}H-^{15}N-HSQC spectra. Due to the high R_2 values in the tandem WW domain constructs, rotational correlation times (τ_m) using R_2/R_1 could not be obtained and relative orientations of the WW domains and so restrictions on the movement of the linker could not be estimated by ^{15}N relaxation.

The $\{^{1}H\}$-^{15}N-NOE data (Figure 4.16d, 4.16g) suggest that the individual WW domains within WW1-WW2 exhibit restricted movement, with the backbone NH bond vectors on the ps to ns timeframe showing values above 0.7. This is consistent with the rigid folded structure of WW domains. Conversely, a high degree of flexibility within the inter-domain linker of both YAP and TAZ is evident. Values obtained from $\{^{1}H\}$-^{15}N-NOE experiments for YAP linker residues were close to zero, and for TAZ were negative, with average values of 0.064 and -0.207, respectively. The missing NOE values in the TAZ linker arise from a lack of NH assignments. The location of YAP and TAZ WW domain linker residue peaks in the central area of the HSQC gave an initial indication that this region is unstructured. Taken together, the data indicate that the linker contains no residual secondary structural elements.

Relaxation data were acquired on the post-titration (2PPxY) YAP tandem WW domain sample (Figure 4.17). The larger R_1 values and more uniform distribution of R_2 relaxation rates indicate the existence of a more stable WW1 conformation in the presence of peptide. This is consistent with the increased intensity of WW1 ^{1}H-^{15}N-HSQC peaks in the proton chemical shift region of 9-10 ppm during the titration experiment, which also suggests enhanced stability of the domain. The backbone dynamics of YAP WW1-WW2 are remarkably similar in both bound and unbound states (Figure 4.17c). The two WW domains remain rigid and the linker retains the high level of flexibility in the presence of 2PPxY peptide.

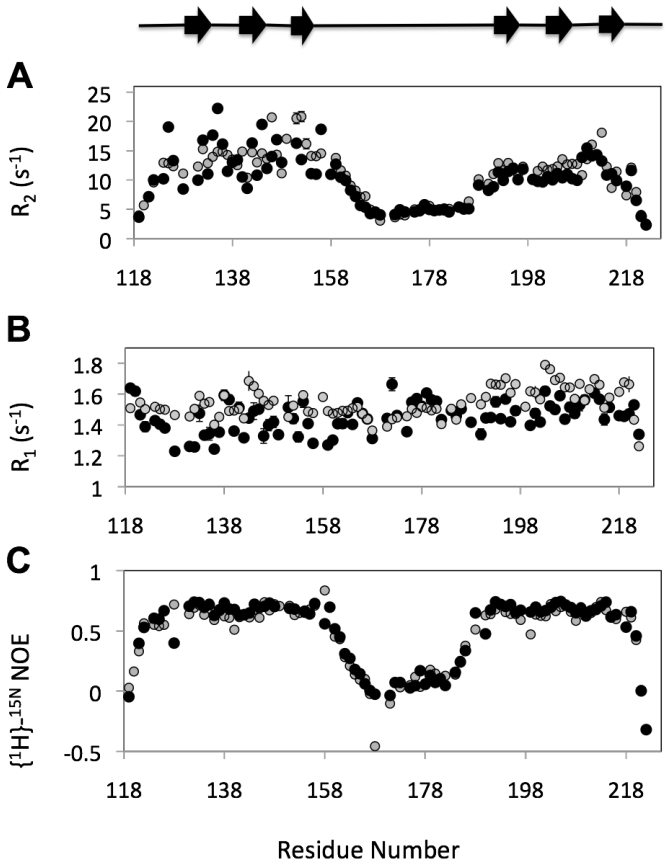


Figure 4.17 YAP WW1-WW2 (A, B) relaxation rates and (C) backbone dynamics: in the absence (black) and presence of dual PPxY peptide (grey).

4.4 Discussion

4.4.1 Identification of TAZ2

In all previously published studies of TAZ, only a single WW domain isoform containing WW1 has been reported (TAZ1). Both the single and tandem WW domain isoforms of YAP, YAP1 and YAP2 had been identified by the mid '90s (Sudol 1994; Sudol et al. 1995). A tandem WW domain isoform of TAZ was identified in medaka fish during the course of this project. YAP has been more extensively studied than TAZ; therefore it is not inconceivable that TAZ2 exists in vertebrate organisms other than medaka but has not yet been found. Differences in the biological roles of YAP1 and YAP2 are currently unclear.

It was previously thought that YAP1 might be anti-apoptotic and YAP2 pro-apoptotic but more recent evidence suggests that the biological activity is dependent on cellular context. The extra WW domain might act to strengthen interactions with a partner protein containing two PPxY motifs, or alternatively recruit two molecules of a

ligand. YAP2 exhibits stronger binding and increased transcriptional activation with the C-terminal fragment of the EGF protein tyrosine kinase ErbB-4 than YAP1 (Komuro et al. 2003). In the same study, YAP2 was shown to be more prevalent than YAP1 to the extent that YAP1 could not be detected, perhaps indicating a context-dependent spatiotemporal requirement for YAP2 rather than YAP1.

4.4.2 Structural analysis of YAP and TAZ WW domains

The structure of human YAP WW1 in complex with the GTPPPPYTVG peptide was first published in 1996 (Macias et al. 1996). The structure of medaka YAP WW2 was determined as part of this project to confirm that the domain contains the three-stranded twisted β-sheet fold typical of other WW domains including YAP WW1. YAP WW2 does fit this description. As observed in WW1, the regions N- and C- terminal to the folded core wrap around the opposite face to the peptide-binding site and interact with a cluster of hydrophobic side chains. Comparison of the rigid region of the YAP WW2 structure (G191-W214) with a range of other group I WW domains (Figure 4.18) yields RMSD values of around 1.5 Å (calculated using Pymol). The most inconsistent area between the WW domains is the flexible loop that links β-strands 1 and 2. This loop is known to tailor the specificity of WW domain binding to partner proteins (Verdecia et al. 2000).

Figure 4.18 Comparison of medaka WW2 (green) with human YAP WW1-N-(n-octyl)-GPPPY-NH2 (cyan), human YAP WW1 L30K mutant-GTPPPPYTVG (magenta), murine-Sav1 WW1 (orange), human WWOX WW1 (blue), and human Magi1 WW1 (purple). The image was made and superimposition carried out using PyMol (MacPyMOL 2009-2010).

Chemical shift-based secondary structure prediction of the TAZ WW domains indicated that they too contain the classical WW domain fold. This is not particularly surprising given the similarities between YAP and TAZ WW domain HSQC spectra. Interpretation of the $\{^{1}H\}$-^{15}N NOE data confirm that YAP and TAZ WW1 and WW2 domains within WW1-WW2 are rigid with average values above 0.7, which indicate restricted movement in the NH bond vectors and are consistent with the presence of secondary structural elements.

4.4.3 Conformational stability of YAP WW domains

Loss of the unstructured region N-terminal to YAP WW1 destabilises the WW domain fold as reported previously (Macias et al. 1996), leading to HSQC spectra representative of a disordered or unstructured polypeptide. This is likely to be an effect of losing a portion of the protein that, although unstructured, is usually in close proximity to the hydrophobic non-binding site of the domain and that forms one half of the 'buckle' that loops around the domain (Figure 1.9a). However, the exact mechanism of this stabilisation is unclear and the Ile is not conserved within WW domains from other proteins, or even YAP WW2 so it seems to be a specific requirement of YAP WW1.

All published structures of YAP WW1 have involved interactions with PPxY containing ligands (Macias et al. 1996; Pires et al. 2001; Aragon et al. 2011). The lack of an unbound YAP WW1 structure could arise from conformational instability of the domain. Another early indicator in this study that WW1 is not as stable as WW2 was the difficulty experienced when assigning WW1 backbone peaks in the WW1-WW2 ^{1}H-^{15}N-HSQC spectrum, which meant that triple resonance data and assignments for isolated WW1 were required to assign equivalent WW1 peaks in the tandem domain. Many cross-peaks for isolated WW1 were weak but detectable whereas the corresponding WW1 signals in the tandem WW domain were not detectable. This is at least partly due to the smaller size of isolated WW1, which is able to undergo faster tumbling than WW1-WW2, leading to the formation of sharper peaks.

Conformational exchange in the non-binding face of YAP WW1 was observed during analysis of ^{15}N relaxation dispersion and R_2 data (Figure 4.15). YAP WW2 is not subject to chemical exchange on the μs-ms timescale (Figure 4.15). A comparison of YAP WW1 and WW2 primary sequences reveals a difference that may potentially contribute to stabilising WW2. In WW2 two hydrophobic residues (I196 and I202) are

located either side of loop 1, adjacent to the hydrophobic core, as compared to the positively charged residues in corresponding WW1 locations (K136 and R142) (Figure 4.19). It is possible that the conformational exchange in WW1 represents equilibrium between folded and unfolded/partially unfolded states. This theory is supported by the fact that the residues involved in chemical exchange contribute to the structural integrity and stability of the domain as part of its hydrophobic core.

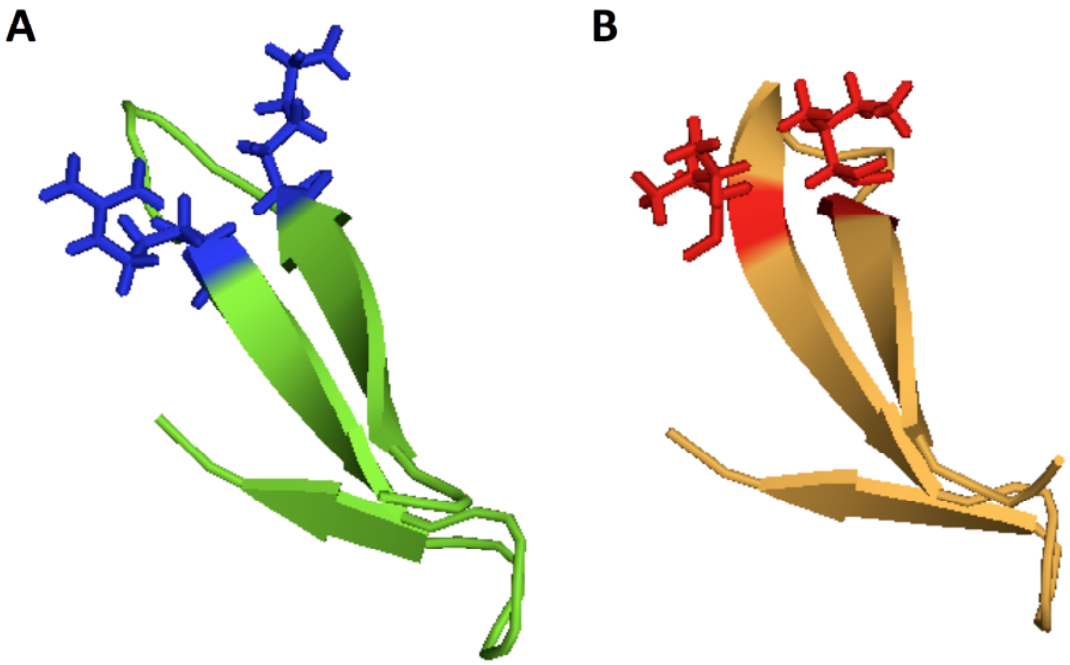


Figure 4.19 Medaka fish YAP (A) (peptide-bound) WW1 (PDB ID: 1K9R Pires et al (2001)), and (B) WW2. (A) The positively charged K136 and R142 residues of WW1 that could repel each other and destabilise the core are highlighted in blue. (B) The hydrophobic I196 and I202 of WW2 that could enhance the stability of the hydrophobic core are in red. I121 is not shown in this image. The figure was made using PyMol (MacPyMOL 2009-2010).

4.4.4 Conformational instability of TAZ WW domains

TAZ WW1-WW2 (Figure 4.11) exhibited more heterogeneous and overall weaker ^{1}H-^{15}N-HSQC signal intensities than YAP WW1-WW2 (Figure 4.7). This is consistent with the elevated ^{15}N R_2 values seen in both TAZ WW domains, which are due to chemical exchange. The chemical exchange would suggest that, as for YAP WW1, both TAZ WW domains are in equilibrium between folded and unfolded/partially folded states. Differences in both R_1 and R_2 indicate that overall molecular motion is greater in TAZ WW1-WW2 than in YAP WW1-WW2. Whilst the difference in R_2 could be

explained by the higher temperature (27.5°C) used for TAZ data acquisition than for YAP (18°C), the difference in R_1 can only be partially ascribed to temperature. The additional dynamics that occur in TAZ WW1-WW2 as compared to YAP WW1-WW2 could be attributed to more marked motions of the two domains, perhaps due to the longer linker, in combination with conformational exchange of the WW domains.

4.4.5 Lack of structure and rigidity in the inter-WW linker

Chemical shift and $\{^1H\}$-^{15}N NOE data suggest that the inter-WW domain linker in both YAP and TAZ lacks secondary structure and is extremely flexible with average $\{^1H\}$-^{15}N NOE values close to or below 0. Negative values reflect fully disordered regions with no restriction to internal motion. Tandem WW domains from several other proteins have previously been studied, some of which also contain linkers without secondary structure (Su(dx) (Fedoroff et al. 2004), FBP21 (Huang et al. 2009), Smurf2 (Chong et al. 2010)) and one that has a rigid α-helical linker that restricts the orientation of the WW domains (Prp40 (Wiesner et al. 2002)).

The unstructured inter-WW domain linkers of Su(dx), FBP21 and Smurf2 are shorter than the YAP and TAZ linkers (35 and 45 residues) at 20, 12, and 20 residues in length. Although these three tandem WW domain linkers are unstructured, they are all subject to some degree of restriction with regard to inter-domain movement and orientation, as determined by $\{^1H\}$-^{15}N NOE data, the presence of medium range NOEs, RDC data, and R_2/R_1 ratios that indicated the two WW domains tumble together in solution (Fedoroff et al. 2004; Huang et al. 2009).

The YAP linker has an average $\{^1H\}$-^{15}N NOE value of 0.064 suggesting little or no restriction on the movement of NH bond vectors. No medium range NOEs were detected within the inter-domain region indicating that YAP WW domains have more freedom of movement than the aforementioned proteins. The TAZ linker is more disordered and flexible than YAP with negative $\{^1H\}$-^{15}N NOE values, which is probably due to the extra 10 residues; this supports the statement made in section 4.4.4 regarding a higher degree of movement between the two domains than in YAP. Although we acquired R_2 and R_1 values from the dynamics data, the conformational exchange that most likely caused the large R_2 values meant that an estimation of motional independence of the WW domains, as was done for FBP21, was not possible.

4.4.6 Conformational significance of peptide binding

Relaxation and backbone dynamics data were acquired for YAP WW1-WW2 bound to dual-PPxY peptide. The R_1 and R_2 data for the peptide-bound form suggest that WW1 exists in a more stable conformation than the unbound form; this is in agreement with ^{1}H-^{15}N-HSQC titration spectra that show increased peak intensities in the 9-10 ppm region and previous observations that YAP WW1 is less stable than other WW domains (Ibragimova et al. 1999).

A recently published study of YAP WW1-WW2 interactions with a phospho-Ser-PPxY peptide derived from Smad1 revealed that WW1 preferentially binds the phospho-Ser and WW2 binds the PPxY motif (Aragon et al. 2011). A comparison of four YAP WW1 domains in complex with either PPxY (Pires et al. 2001) or phospho-Ser (Aragon et al. 2011) peptides revealed RMSD values of more than 2 Å; clear differences in domain conformation and peptide orientation can be seen (Figure 4.20). The YAP-phospho-Ser WW1 structure is in a wider, more open conformation than the YAP-PPxY WW1 structure. There is also a noticeable difference in the specificity loop between the β1 and β2 strands. As WW1 can bind two different motifs and WW2 is only known to bind one motif, it is possible that the conformational instability of WW1 contributes to the altered specificity.

No significant differences in the backbone dynamics of the bound and unbound forms of YAP WW1-WW2 were observed. In the peptide-bound form the linker still displayed all the hallmarks of an unstructured, flexible region. The solution structure of YAP WW1-WW2 in complex with the Smad1 phospho-Ser-PPxY peptide revealed an inter-domain linker with a helix-loop-helix structure (Aragon et al. 2011). The tandem WW domain structure was not deposited in the PDB and the residue numbers are not disclosed, therefore it is difficult to make a comparison with our data. The $\{^{1}H\}$-^{15}N NOE values for the YAP and TAZ WW1-WW2 gradually approach 0 over approximately ten residues at either end of the linker, which could account for the presence of potential helices flanking the unstructured flexible region and the region N-terminal to WW2 also appeared to contain a potential helix/coil (Figure 4.10).

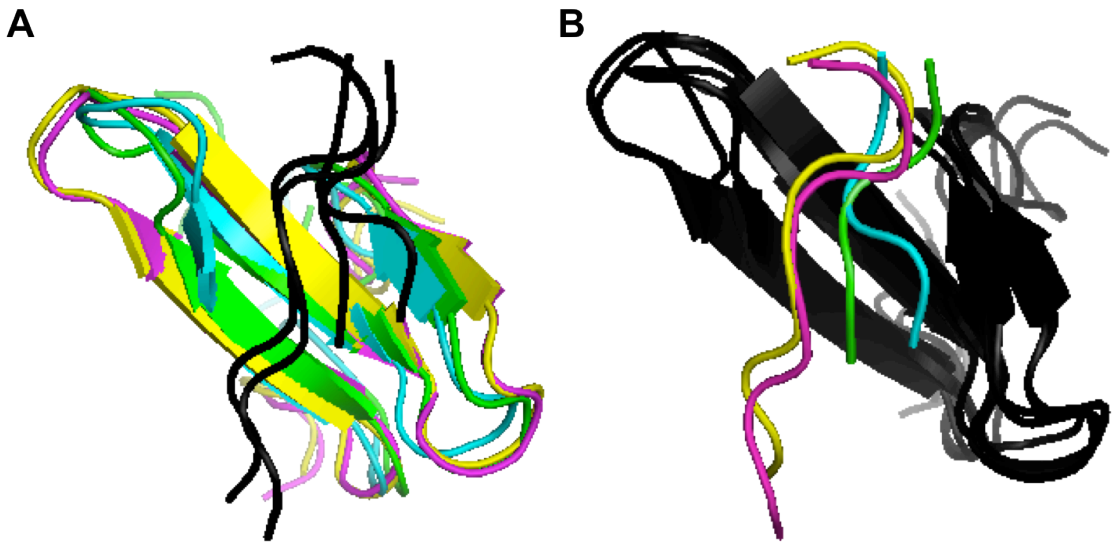


Figure 4.20 YAP WW1 solution structures in complex with various peptide ligands. (A) WW domain conformations and (B) peptide orientations are illustrated in colour. Green represents WW1-N-octyl-GPPPY-NH2 (PDB ID: 1K9Q); cyan represents WW1-acetyl-PLPPY (PDB ID: 1K9R); magenta represents WW1-diphospho-Ser (PDB ID: 2LAX); yellow represents WW1-monophospho-Ser (PDB ID: 2LAY).

4.4.7 Mechanism of WW1-WW2 interaction with PPxY peptides

Titrations of YAP and TAZ WW1-WW2 with single and dual PPxY peptides (GTPPPPYTVG and GTPPPPYTVGGGGGGGTPPPPYTVG) revealed that both WW domains of both proteins bind to both ligands. The pattern of chemical shift changes is very similar between YAP and TAZ, and also between the titrations with the two peptides (Figure 4.14). At each titration point, peaks from both WW1 and WW2 residues moved consistently and when represented graphically the shapes of the curves were similar. Therefore it is highly likely that peptide binding to the two WW domains within WW1-WW2 occurs simultaneously, not sequentially, and that the affinity of WW domains for peptide are similar.

Titrations with single PPxY peptide. The calculated K_d values for the two domains in each protein construct for single PPxY peptide were not vastly different, in the range of 273 ± 116 μM to 330 ± 116 μM. In both YAP and TAZ WW1 peaks exhibited significantly larger shifts overall than WW2 residues; the similarity of K_d values suggests that the larger effect of ligand binding upon the chemical environment of WW1 nuclei is not due to higher affinity. All evidence suggests that the unbound WW2

population exists in one stable conformation, whereas WW1 samples at least two conformations. The WW2 conformation appears to undergo minor structural alterations during interactions with PPxY peptide. One or more of the multiple WW1 conformations may be less similar to the bound state and undergo larger binding-induced conformational changes to produce one stable peptide bound conformation. Binding to peptide would thus have a larger impact on the chemical environment of WW1 than WW2 nuclei.

Implications of tandem WW domains. That YAP and TAZ both contain a WW domain repeat might imply the presence of a second domain confers an advantage over the single WW isoform, or alternatively the second WW domain might be a redundant feature. The latter is less likely as the second WW domain is present in both proteins, at least in medaka fish. As the two WW domains in WW1-WW2 bind the same partner simultaneously and with equal affinity, and also the inter-domain linker is flexible and long enough to allow for various geometries, YAP and TAZ may be able to bind two separate partner molecules or two PPxY motifs on one target protein. Binding two proteins might be prevented due to steric hindrance, unless the two partner molecules also interacted with each other and formed a trimeric complex. It is quite possible that YAP/TAZ WW1-WW2 could interact with two binding motifs on a single partner molecule if the sites were in close enough proximity. Such binding would presumably increase the strength and specificity of the interaction.

Previous studies have noted more than a doubling in affinity of tandem WW domains for certain binding partner molecules and in many of these cases the domains are thought to act cooperatively. The affinity of Smurf WW2-WW3 for a Smad single PPxY peptide is significantly higher than that of YAP/TAZ WW1-WW2 for single PPxY, with a K_d value of 1.7 ± 0.4 μM as compared to ~300 μM. In this Smurf-Smad scenario, Smurf WW2 alone cannot bind peptide but the tandem WW Smurf-Smad peptide interaction is enhanced relative to WW3 alone (Chong et al. 2010). In the case of the group III tandem WW domains of FBP21, a 30-60 fold increase in affinity was observed for peptide containing two proline-rich motifs in comparison to peptide containing one binding motif (Huang et al. 2009). The initial affinity of FBP21 tandem WW domains for a single-PPR binding motif was, however, much lower than we have calculated for YAP and TAZ WW1-WW2 for single PPxY peptide, with FBP21-single-PPR peptide K_d values of 1-3 mM compared to around 300 μM for YAP/TAZ.

Titrations with dual PPxY peptide. During the titrations of YAP and TAZ with dual PPxY peptide, peak movements were observed at much lower peptide/protein ratios (5-10 times) than in the single PPxY experiments. TAZ binds the dual PPxY peptide with higher affinity (48.5 ± 12 μM) than single PPxY peptide. Based on the chemical shift changes, YAP affinity with the dual PPxY peptide was estimated to be similar to that of TAZ. The K_d could not be calculated for YAP with confidence; typically the lower K_d calculation limit using NMR is 10-100 μM (Fielding 2007). The TAZ and YAP dual-PPxY interaction is close to or perhaps beyond the sensitivity limit of NMR spectroscopy. Unlike the FBP21 tandem WW-dual PPR peptide binding, the 6-fold enhancement of YAP and TAZ WW1-WW2 affinity with dual PPxY peptide over single PPxY peptide is not likely to be attributed to inter-WW cooperativity. Ideally other methods of affinity determination would have been employed, but unfortunately fluorescence spectroscopy was unsuccessful and access to large amounts of peptide and ITC equipment were unavailable. The linker residue backbone NH chemical shifts were relatively unaffected during titration with dual PPxY peptide (Figure 4.13) and, consistent with this, little effect on $\{^1H\}$-^{15}N NOE values was observed (Figure 4.16).

Effect of peptide interactions on inter-WW linker. Chemical shift changes within the WW domain and linker residues were similar in both dual and single PPxY titrations. The evidence indicates that the linker remains largely unstructured and flexible; therefore this particular dual PPxY peptide does not induce allosteric conformational change in YAP and TAZ WW1-WW2. An avidity-based mechanism in which binding of one PPxY motif to one WW domain brings the other PPxY motif into proximity to the other WW domain, thereby limiting the spatial search required for the second WW-PPxY binding event to occur is a more probable explanation for the > 2-fold increase in affinity. Based on the small effects on linker residues and the lower K_d of dual PPxY binding, it is possible that the two domain-bound form inter-converts between one domain-bound forms.

Implications of dual PPxY peptide interactions with tandem WW domains. Enhanced dual PPxY binding by tandem YAP and TAZ WW domains over single PPxY peptide has been clearly established. Such enhanced affinity of dual PPxY binding may be important in YAP2 or TAZ2 interaction with binding partners such as LATS1, ErbB4 and AMOT that contain multiple PPxY motifs. The two PPxY motifs in LATS1, a kinase that is part of the core Hippo pathway cassette and that phosphorylates YAP and

TAZ, are separated by about 180 amino acids but it is possible that these PPxY motifs are spatially close enough to bind to tandem WW domains in a single molecule of YAP2 or TAZ2. Evidence suggests that both LATS1 PPxY motifs and both YAP WWs are necessary for full strength YAP-LATS1 binding as determined by transcriptional reporter assays (Oka et al. 2008). In the case of ErbB4, in which two of the PPxY motifs are seventeen amino acids apart, YAP2 (two WW domains) co-activates ErbB4-regulated genes more strongly than YAP1 (one WW domain) (Komuro et al. 2003). Recently the YAP tandem WW domain was found to bind a Smad1 peptide containing a phospho-Ser and a single PPxY motif (Aragon et al. 2011); the presence of both domains of YAP was necessary for the interaction, further compounding the importance of WW domain repeats in enhancing interactions and specificity.

CHAPTER 5: WW DOMAINS OF THE UPSTREAM HIPPO PATHWAY PROTEINS SALVADOR AND KIBRA

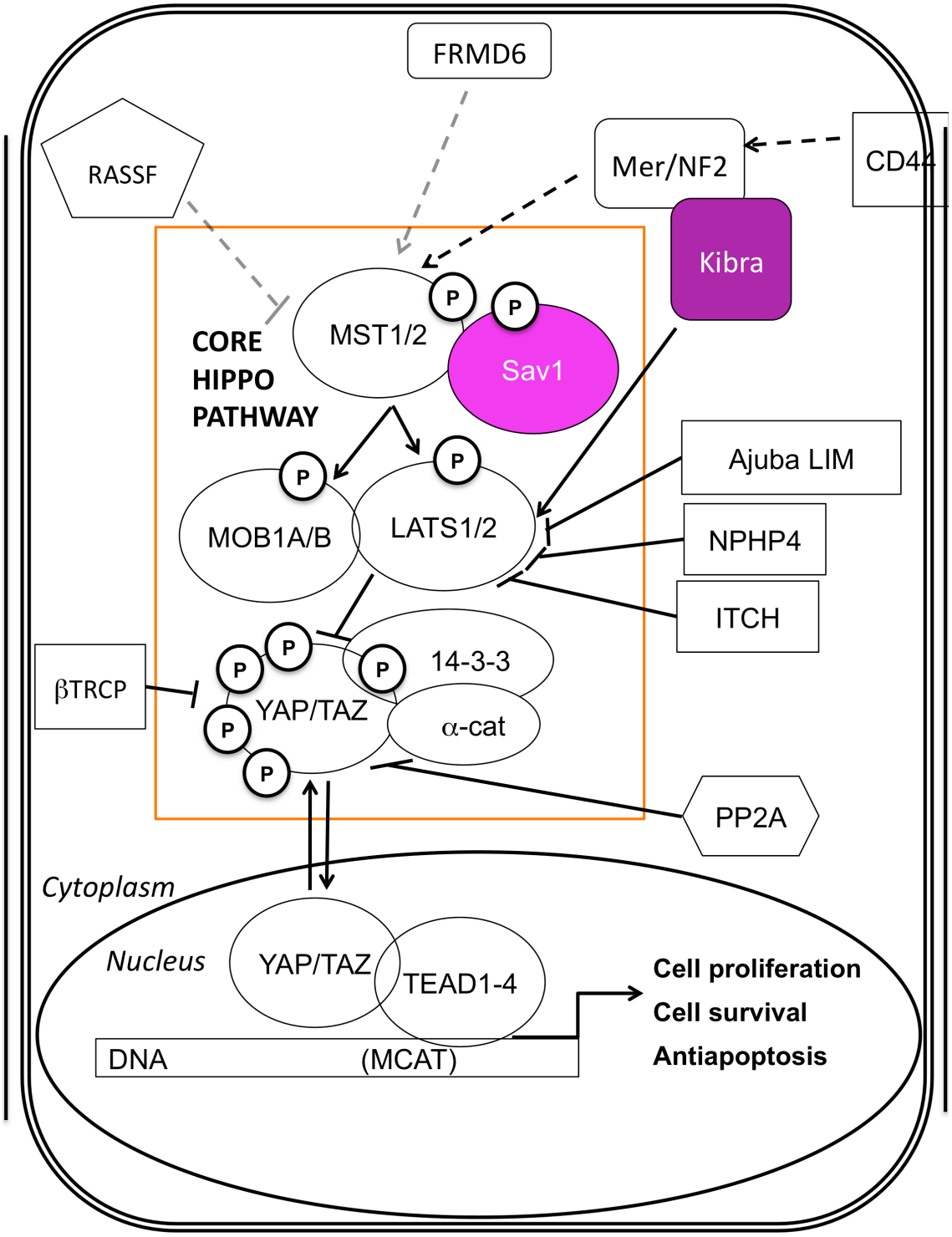


Figure 5.1 Location of Sav (magenta) and Kibra (purple) within the Hippo pathway. Core Hippo pathway components are located within the orange box.

5.1 Introduction

Sav and Kibra are both Hippo pathway proteins (Figure 5.1) that contain a pair of WW domains (Figure 5.2). There is no inter-WW linker in these proteins, unlike in the tandem WW domains of YAP and TAZ (which have linker lengths of 35 and 45 residues, respectively). Both Sav and Kibra also contain an atypical WW domain whereby non-Trp residues are present at the site of the second conserved Trp (Figure 5.3). Mouse Sav WW2 forms a homodimer (Ohnishi et al. 2007). Human Sav WW2 has the same primary structure as mouse Sav WW2, but *Drosophila* Sav WW2 differs (Figure 5.3). To date, no structural studies of Kibra WW domains have been published. Kibra contains a Trp C-terminal to the WW domain that may or may not affect the structure or function of the domain.

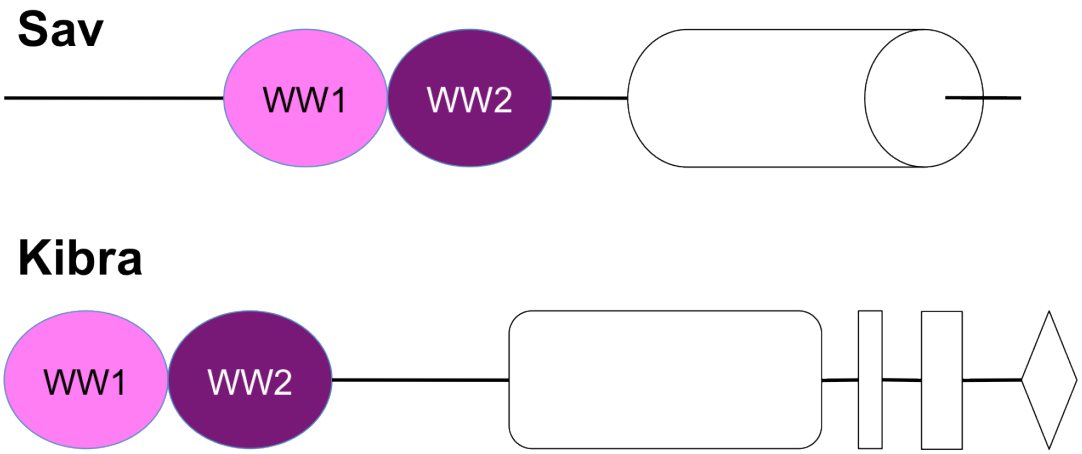


Figure 5.2 Modular domains of Sav and Kibra. The typical WW1 domains are shown in pink and the atypical WW2 (or 'Wx') domains are in purple. Domains and linker regions are to relative scale. Domains that are not studied in this chapter are unlabelled and uncoloured.

```
Hu Kibra WW2          LPLGWEEAYDPQVGDYFIDHNT-KTTQIEDP
Dros Sav WW2          LPVGWRRVVSKMHGTYYENQYT-GQSQRQHP
Hu Sav1 WW2           LPPGWERVESSEFGTYYVDHTN-KKAQYRHP
Hu Magi1 WW2          LPAGWEKIEDPVYGIYYVDHIN-RKTQYENP
Hu WWOX WW2           LPY-WEQETDENGQVFFVDHIN-KRTTYLDP
Hu Smurf1 WW2         LPPGWEIRNTATGRVYFVDHNN-RTTQFTDP
Dros Su(dx) WW4       LPPGWEIRYTAAGERFFVDHNTRRTTTFEDP
                      **  *.         :: :: .   :   .*
```

Figure 5.3 Multiple sequence alignment of atypical WW domains (made using ClustalW2). The locations of the first conserved Trp residue and replacement of the second conserved Trp are highlighted in red.

The aims of the work in this chapter were to characterise the tandem WW domains from the Hippo pathway proteins Sav and Kibra using NMR spectroscopy. Recent collaborators provided the cDNA plasmids for these proteins (N. Tapon and J. Kremerskothen). These collaborators do not work with medaka models and so the proteins are from different organisms (*Drosophila melanogaster* and human) to the previous work. The results presented here provide hints as to the stability and binding capabilities of the WW domains from *Drosophila* Sav and human Kibra but there are still many unanswered questions regarding the structure and function of the tandem WW domains from these proteins.

5.2 Materials and methods

5.2.1 Cloning of DNA fragments

Using the Sav and Kibra cDNA plasmids as template for PCR, WW domain containing DNA fragments were amplified using Phusion DNA polymerase and cloned into Vanderbilt vectors as described in Chapter 2. Primers and constructs can be found in Appendix I. Constructs were designed so that the equivalent residues to I121 of YAP were present (which is not conserved throughout WW domains) to provide a sufficient length of peptide to flank the WW domain core

5.2.2 Expression and purification of Sav WW domain protein constructs

Drosophila *Sav WW1 and WW2 expression and purification. Drosophila* Sav WW1 and WW2 domains were expressed using the pSV281 and pBG101 Vanderbilt vectors, respectively, in Arctic Express (RIL) chemically competent cells. The cells were grown to OD_{600} of 0.6-0.9 at 37°C and induced overnight at 12°C using 0.5 mM IPTG. ^{15}N-labelled preparations were inoculated as described in Chapter 2, with the exception of the primary culture, which was always started from an agar plate rather than a glycerol stock. Cells were harvested by centrifugation and the cell pellet was resuspended in His-buffer A with one 'Complete mini protease inhibitor cocktail' tablet/50 ml buffer. Protein was purified using nickel affinity chromatography as described in Chapter 2.

The His-GST tag was cleaved overnight from WW2 using 3C protease in approximately 5 ml buffer (50 mM Tris pH 7.5, 150 mM NaCl, 2% glycerol, and 1mM DTT) rocking at 4°C. 5 ml His-buffer A was added to the post-cut protein, which was

separated from the tag using a 5 ml HisTrap FF column (GE Healthcare) with an elution gradient of at least 16 column volumes up to a concentration of 0.5 M imidazole. Cut WW2 was present in the flow-through and the first two elution-fractions. The effectiveness of the protease and purity of the WW domains were verified by SDS-PAGE as described in Chapter 2.

Purification of human and Drosophila *Salvador WW1-WW2 tandem domain constructs.* Tandem WW domains from Salvador were expressed in the same manner as the individual WW domains. The only deviation from the individual Salvador WW domain protocol was the buffer content. A solubility screen found Tris-buffers and high salt concentrations to be unfavourable. Therefore, the His-buffers were replaced with 50 mM HEPES pH 7.5 and 100 mM NaCl, with 20 mM (buffer A) and 500 mM (buffer B) imidazole for the initial purification. If a GST-tag was to be removed (i.e. when using protein from the pBG101 vector), the WW1-WW2 was separated from the tag using a GST-trap FF column.

5.2.3 Expression and purification of human Kibra WW domain protein constructs

Joachim Kremerskothen (Universitat Münster) sent individual and tandem Kibra WW domain plasmids, which were transformed into BL21 (DE3) cells but protein expression and purification trials were unsuccessful (proteins either did not express or were insoluble). Both short and long Kibra tandem WW domain constructs were re-cloned into the Vanderbilt vector pSV281 (kan^R) and expressed in BL21 (DE3) cells at 37°C for 4 hours. The long construct contains an extra Trp residue C-terminal to the core atypical WW2 domain. ^{15}N-labelled cultures were inoculated as described in Chapter 2. The proteins were purified by nickel ion chelating chromatography as previously described (Chapter 2).

5.2.4 Glutaraldehyde cross-linking experiment with Drosophila *Sav WW domains*

The pure proteins (WW1, WW2 and WW1-WW2) were exchanged into PBS using a PD10 desalting column and incubated with a final concentration of 0.1% v/v glutaraldehyde. The incubations took place at room temperature for 0, 1 and 30 min, and 4°C for 0, 5 and 30 min. The 20 µl reactions were terminated by the addition of 2 µl 1 M Tris-HCl pH 7.5. The samples were analysed by SDS-PAGE as described in Chapter 2.

5.2.5 2D NMR conditions and experiments with Sav WW domains

Sample preparation. Pure ^{15}N-labelled *Drosophila* Sav WW1 and WW2 without tags were buffer exchanged (into 50 mM MES pH 6.5, 50 mM NaCl) by repeated dilution and concentration as previously described. 10% D_2O was added to the final samples which were transferred to standard NMR tubes. Several buffers were used for the human and *Drosophila* WW1-WW2 proteins in an attempt to improve the HSQC spectra (Table 5.1).

Table 5.1 Buffer conditions used for Sav WW1-WW2 NMR experiments

Construct	**Buffer**	**Salt**	**Additives**	**pH**
Drosophila	50 mM MES	50 mM NaCl		6.5
Drosophila	50 mM MES	150 mM NaCl		6.5
Human	50 mM KH_2PO_4	200 mM NaCl	5 mM EDTA	5.5
Drosophila	50 mM HEPES	50 mM NaCl	1 mM EDTA	6.8

NMR experiments. The experimental parameters for all Salvador WW domains included an optimal temperature of 12°C, nitrogen spectral width of 1640.9 Hz and nitrogen carrier frequency of 120 ppm. Experimental procedures and data processing/analysis were carried out as described in Chapter 2. In addition to varying the buffers and temperatures to obtain the best HSQC spectrum for each construct, experiments were carried out to assess the interaction of the WW domains with each other and with the GTPPPPYTVG peptide.

In order to determine whether the presence of *Drosophila* Sav WW1 affects the secondary structure of Sav WW2, an NMR 'mixing experiment' was conducted, whereby unlabelled lyophilised WW1 was added to a known concentration (~250 μM) of ^{15}N-labelled WW2 incrementally. NH chemical shifts were monitored in ^{1}H-^{15}N-HSQC spectra as described in Chapter 4.

The single PPxY peptide GTPPPPYTVG was dissolved in the relevant NMR buffer and titrated into known concentrations of *Drosophila* Sav WW1, WW2 and WW1-WW2 in order to assess the effects of peptide binding. The titration experiments were monitored as described previously (Chapter 4).

5.2.6 2D NMR conditions and experiments with Kibra tandem WW domains

Both ^{15}N-labelled human Kibra WW1-WW2 proteins (WW1-WW2(s) and WW1-WW2(l)) with His-tag were exchanged into 50% PBS and 10% D_2O. Experimental procedures and data processing/analysis were carried out as described in Chapter 2. The optimal temperature for both tandem WW domain constructs was 20°C, the nitrogen spectral width was 1884 Hz and nitrogen carrier frequency 120 ppm. The GTPPPPYTVG peptide was titrated into ^{15}N-labelled Kibra tandem WW domain proteins as described in Chapter 4.

5.3 Results

5.3.1 Expression and purification of Sav protein constructs

Drosophila Sav WW1 was grown in ^{15}N-labelled M9 medium, and induced, purified (Figure 5.4), and characterised with a His-tag from the pSV281 vector. Fraction '3' shown in Figure 5.4 was used to make the WW1 NMR sample. Fractions 1 and 2 were discarded because they appeared to be of a lower molecular weight and so may have introduced an element of heterogeneity into the sample. Cleaved Salvador WW2 has residual affinity for nickel columns. During the separation of WW2 from the His-GST tag, a long elution gradient (16 column volumes in a 5 ml His-Trap FF column) was used to successfully purify the protein (Figure 5.5).

Optimisation of buffer conditions. Sav WW1-WW2 purification conditions were optimised using micro-drop solubility screens (Lepre et al. 1998). Tris based buffers were avoided, as they increased protein precipitation. The best buffer condition for the initial affinity purification for tandem Sav WW domains was HEPES pH 7.5 with 100 mM NaCl and 20 mM/500 mM imidazole (buffer A/B). For Sav WW1-WW2 pBG101, a glutathione sepharose column (GSTrap FF, GE Healthcare) was used to separate the protein from the His-GST tag (Figure 5.6).

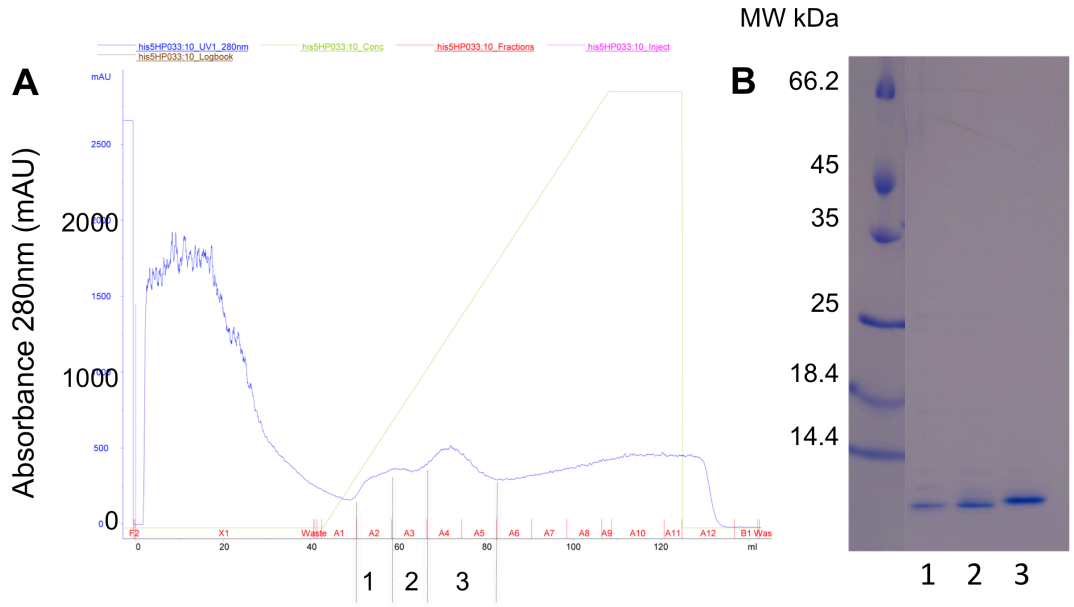


Figure 5.4 *Drosophila* Sav WW1 (A) His-affinity elution profile from a 5 ml His-Trap FF column; the blue line represents the A_{280} and the green line is the imidazole gradient. (B) Coomassie stained SDS-PAGE analysis; 5 µl marker was loaded (on the left with the band MWs labelled) and 10 µl each fraction diluted 1:1 with 2X sample buffer (Appendix III-c) were loaded in lanes 1-3. Fractions 1 and 2 have a slightly lower MW than fraction 3 and may be partially degraded. Fraction 3 was used to make the NMR samples.

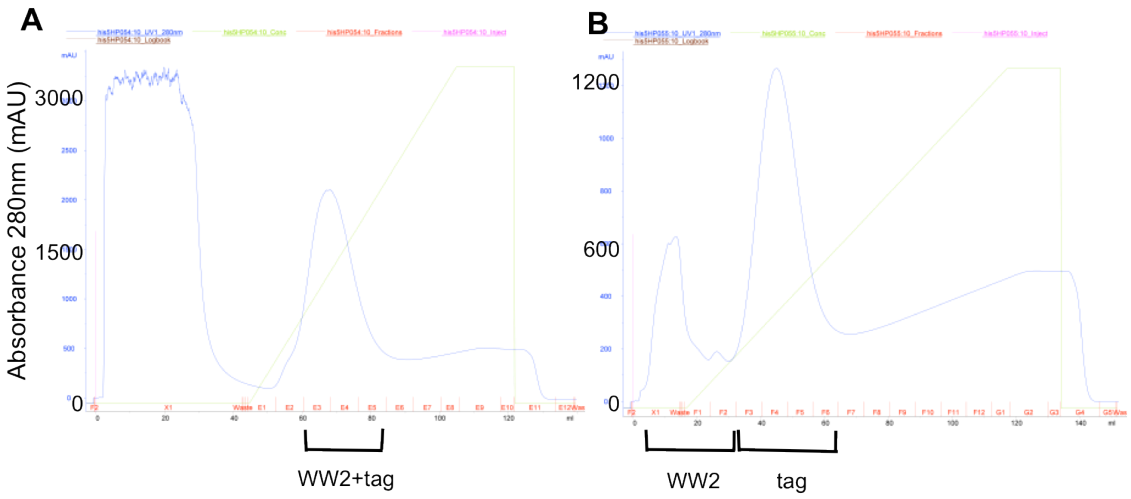


Figure 5.5 (A) The initial and (B) post-cut His-GST-tagged ^{15}N-labelled *Drosophila* Sav WW2 nickel affinity (5 ml His-Trap FF column) purification profiles. the blue line represents the A_{280} and the green line is the imidazole gradient.

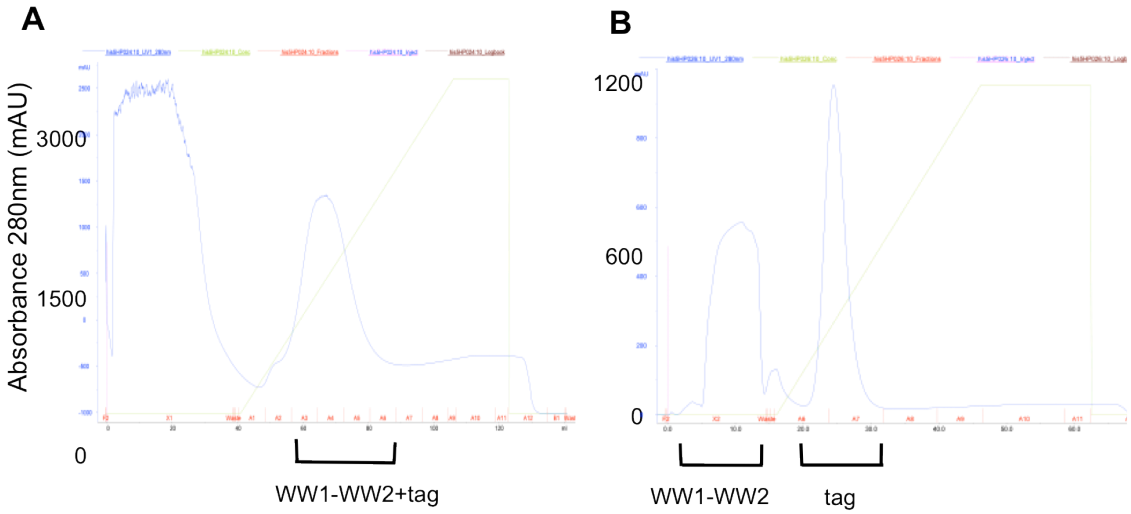


Figure 5.6 ^{15}N-labelled *Drosophila* Sav WW1-WW2 purification; the blue line represents the A_{280} and the green line is the imidazole gradient. (A) The His-GST tagged WW1-WW2 was initially purified using a 5 ml HisTrap FF column (GE Healthcare). (B) 3C protease cleaved WW1-WW2 was separated from the His-GST tag using a 5 ml GSTrapFF column (GE Healthcare).

5.3.2 Glutaraldehyde cross-linking experiment with Sav WW2

A glutaraldehyde cross-linking experiment was conducted in an attempt to discover whether *Drosophila* Sav WW2 dimerises in a similar manner to murine Sav WW2. Glutaraldehyde would covalently link Arg and Lys side chains that were in close proximity to each other. The presence of a covalently linked dimer could then be detected using Coomassie stained SDS-PAGE. Arg and Lys residues are present in the proposed Sav WW2 dimer interface. The result of this glutaraldehyde cross-linking experiment suggests that *Drosophila* Sav WW2 does not dimerise as no increase in the molecular weight of WW2 is observed by SDS-PAGE (Figure 5.7). Another method that could have been utilised to answer this question is gel filtration chromatography (GFC), however the WW domain is very small at around 4 kDa and the GFC columns available had previously undergone some degree of resin packing such that detecting such a doubling in MW would be unlikely. If the domain was much larger than 4 kDa then light scattering methods might also have been used.

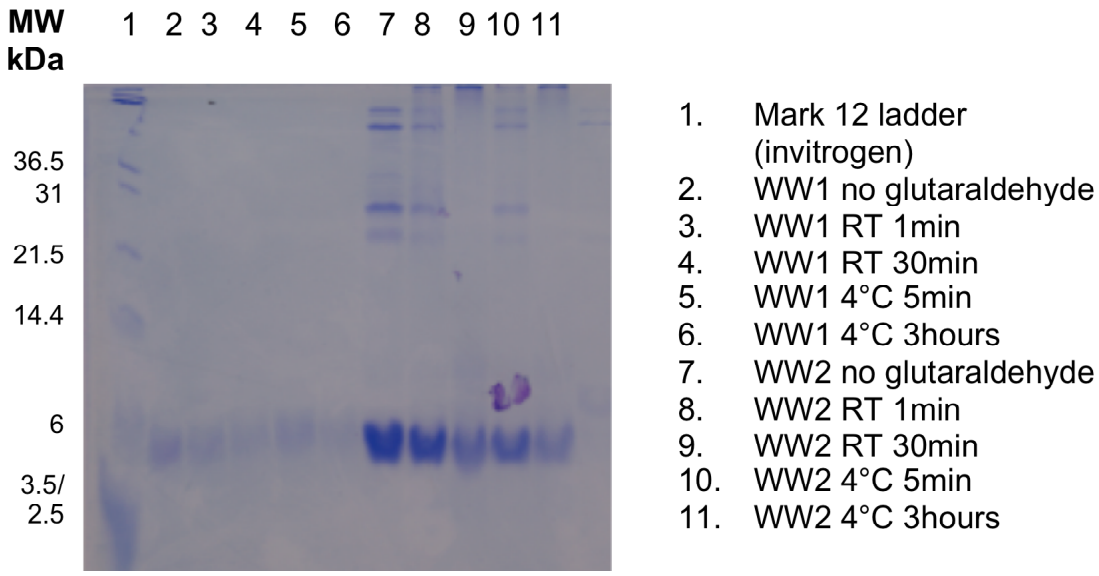


Figure 5.7 Glutaraldehyde cross-linking experiment. *Drosophila* Sav WW1 and WW2 were incubated with 0.1% glutaraldehyde and electrophoresed in a 20% polyacrylamide gel. No increase in molecular weight was observed for either protein fragment.

5.3.3 2D NMR conditions and experiments with Sav WW domains

The *Drosophila* Sav WW1 ^{1}H-^{15}N-HSQC spectrum (Figure 5.8a) exhibits only 11 backbone NH peaks, significantly fewer than the expected 32 non-Pro and non-NH_2 peaks (or 50 non-Pro peaks including His-tag residues). The low number of peaks observed suggests broadening of the line widths and is consistent with a protein in equilibrium between two states. It is probable that WW1 exists in unfolded/partially unfolded and completely folded conformations. WW1 undergoes binding-induced folding upon titration with single PPxY peptide, as evidenced by the appearance of the expected number of well dispersed peaks (Figure 5.8). The PPxY-bound WW1 peaks around 9-10 ppm in the proton dimension disappear as temperatures increase from 12°C to 20°C. An extra indole peak is present around ^{1}H 10 ppm in the peptide bound WW1, which indicates that at least one of the Trp side chains undergoes chemical exchange. The extra indole peak becomes more intense with the increase in temperature from 12°C to 20°C (Figure 5.9).

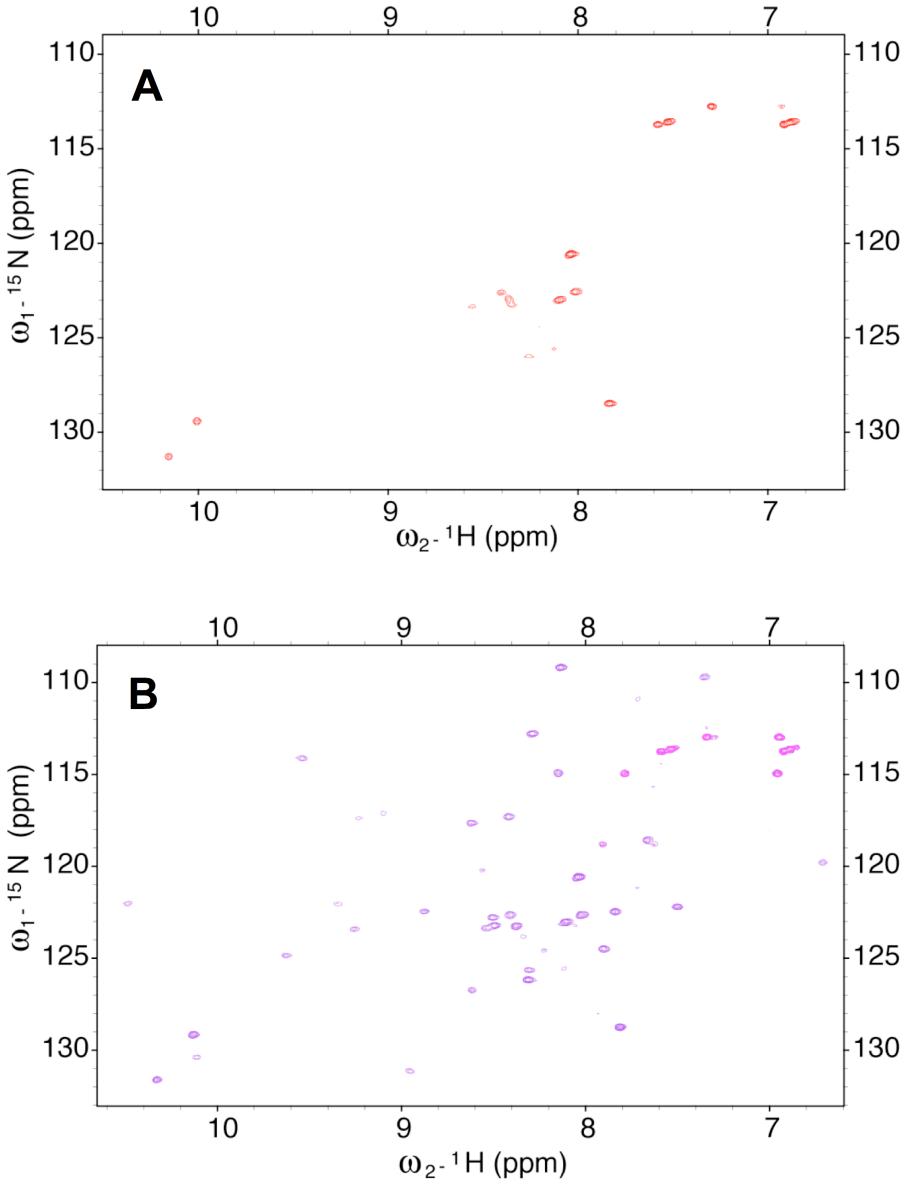


Figure 5.8 *Drosophila* Sav WW1 ^{1}H-^{15}N-HSQC spectra in (A) the absence (red) and (B) the presence (purple) of PPxY peptide at 12°C. Very few peaks are observed in (A) indicating that the domain exists in two states (possibly unfolded and folded). (B) Peptide binding induces conformational stability in *Drosophila* Sav WW1.

The ^{1}H-^{15}N-HSQC spectrum for WW2 (Figure 5.9) suggests the domain is folded but the peaks between 9 and 10 ppm are weaker than the majority of other peaks, as was observed for YAP and TAZ. Like Sav WW1, Sav WW2 is more stable at lower temperatures (Figure 5.9). NMR titrations were conducted on ^{15}N *Drosophila* Sav WW2 with single PPxY peptide and no changes were observed in ^{1}H-^{15}N chemical shifts; therefore WW2 is not capable of binding this PPxY peptide. This result is not

entirely unexpected as an Arg residue replaces the second conserved Trp in WW2 and this Trp normally plays a key role in PPxY binding.

Lyophilised unlabelled Sav WW1 was added to ^{15}N-labelled *Drosophila* Sav WW2 in an NMR mixing experiment in order to assess potential interactions between the two WW domains. No visible changes were observed in the ^{1}H-^{15}N-HSQC spectrum of WW2; therefore it is unlikely that *Drosophila* Sav WW1 and WW2 interact within the tandem WW domain. As Sav WW2 is unaffected by the presence of single PPxY peptide and is also unaffected by WW1, it is unlikely that the tandem WW1-WW2 binds peptide in a cooperative manner.

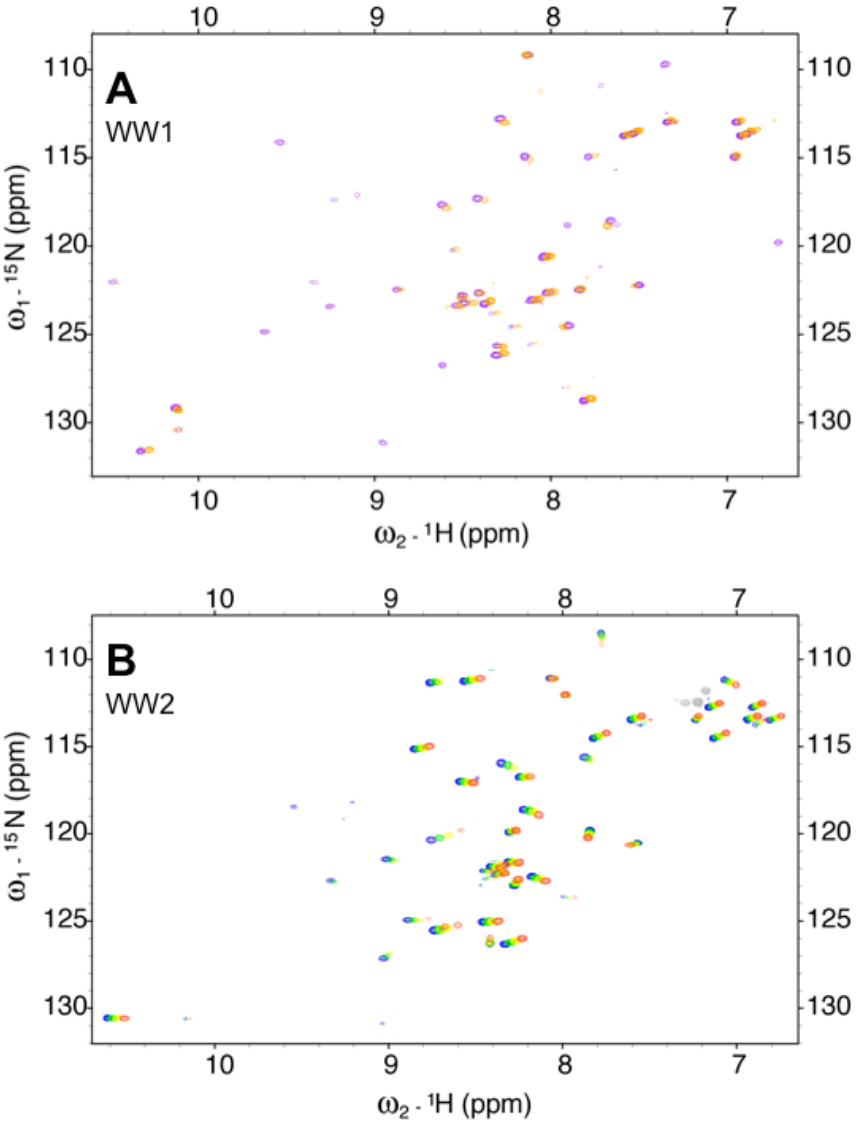


Figure 5.9 *Drosophila* Sav (A) peptide-bound WW1 and (B) native WW2 ^{1}H-^{15}N-HSQC spectra at a range of temperatures. (A) WW1 peaks at 12°C are purple and at 20°C are orange. (B) WW2 peaks are shown in blue at 12°C, green at 16°C, yellow at 20°C and finally in orange at 25°C.

The *Drosophila* Sav WW1-WW2 NMR sample was difficult to make as the protein precipitated easily, particularly upon concentration. The ^{1}H-^{15}N-HSQC spectra (Figures 5.10 and 5.11) display peaks of different intensities, which is a sign of conformational instability. Also there are many small peaks present, particularly around ^{15}N and ^{1}H shifts of 130 ppm and 8 ppm, suggesting proteolytic degradation of the sample. The well dispersed peaks in the region of 9-10 ppm belong to WW2 within the tandem WW domain, as demonstrated by overlaying the spectra (Figure 5.10). The quality of the WW1-WW2 ^{1}H-^{15}N-HSQC spectrum is not improved by varying temperature or titrating the sample with single PPxY peptide, providing further evidence that the labelled protein is subject to degradation. Adding peptide does, however, induce the appearance of weak peaks around 9-10 ppm, which is consistent with the binding-induced folding of WW1 (Figure 5.11). Human Sav WW1-WW2 was also difficult to work with and produced a similar (and perhaps worse) ^{1}H-^{15}N-HSQC to *Drosophila* Sav WW1-WW2.

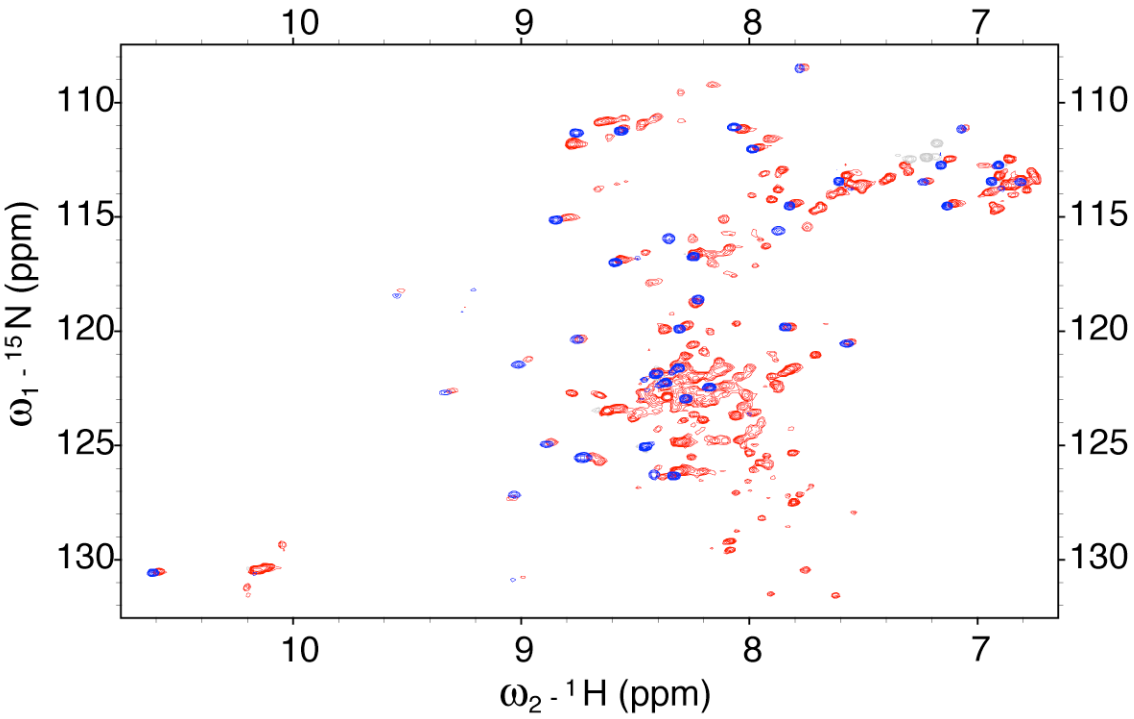


Figure 5.10 ^{1}H-^{15}N-HSQC spectra of *Drosophila* Sav WW2 recorded at 12°C with 8 scans (blue) and WW1-WW2 recorded with 240 scans (red). A longer acquisition time was necessary for WW1-WW2 because of the low protein concentration (~100 μM) as compared with a protein concentration of ~500 μM WW2.

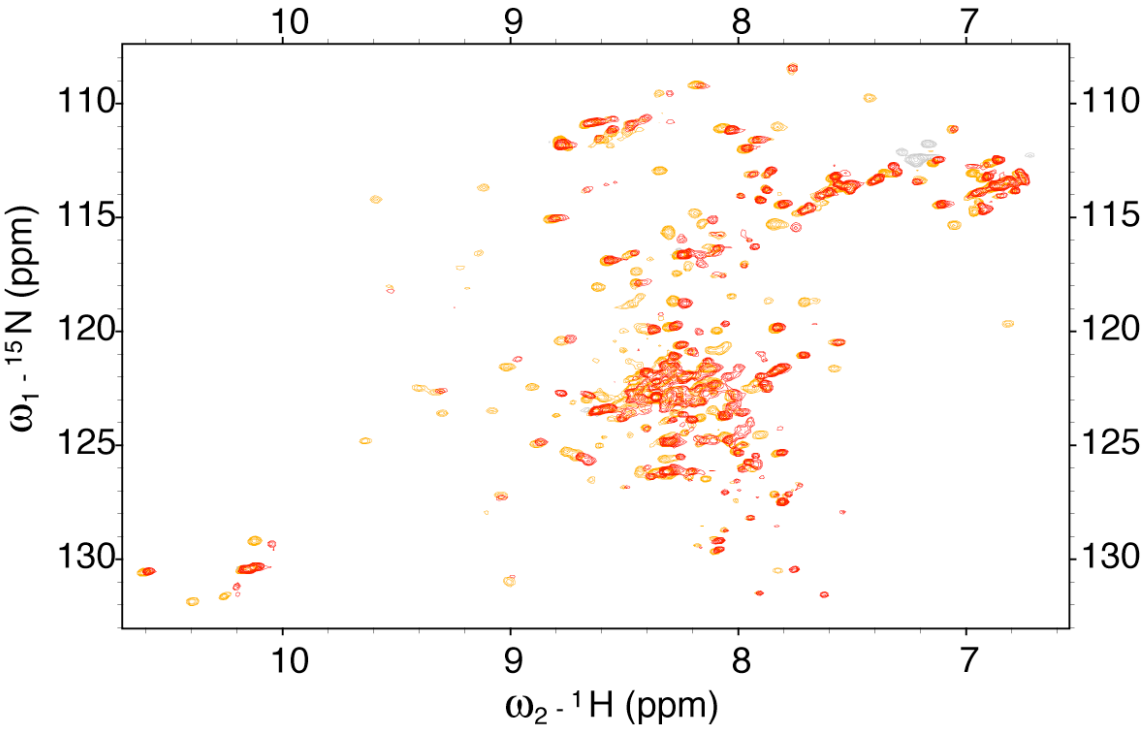


Figure 5.11 Overlaid ^{1}H-^{15}N-HSQC spectra of *Drosophila* Sav WW1-WW2 without (red), and with (orange) 1.PPxY peptide (as was used in Chapter 4). Both spectra were recorded at 12°C with 240 scans (overnight). WW1-WW2 with peptide displayed more peaks in a better-dispersed distribution, but the heterogeneity in the sample was not reduced (i.e. many peaks are still overlapped in the centre of the spectrum and many small degradation peaks are present)

5.3.4 Expression and purification of Kibra WW domain constructs

Expression trials of individual Kibra WW domains (WW1 and WW2) in the vectors sent by Joachim Kremerskothen (Universität Münster) in BL21 (DE3) cells were unsuccessful. Kibra WW1-WW2(s) and WW1-WW2(l) were successfully re-cloned into the pSV281 vector and expressed in BL21 (DE3) cells at 37°C. Kibra tandem WW domains were purified using nickel ion affinity chromatography with a 5 ml His-Trap FF column (Figure 5.12).

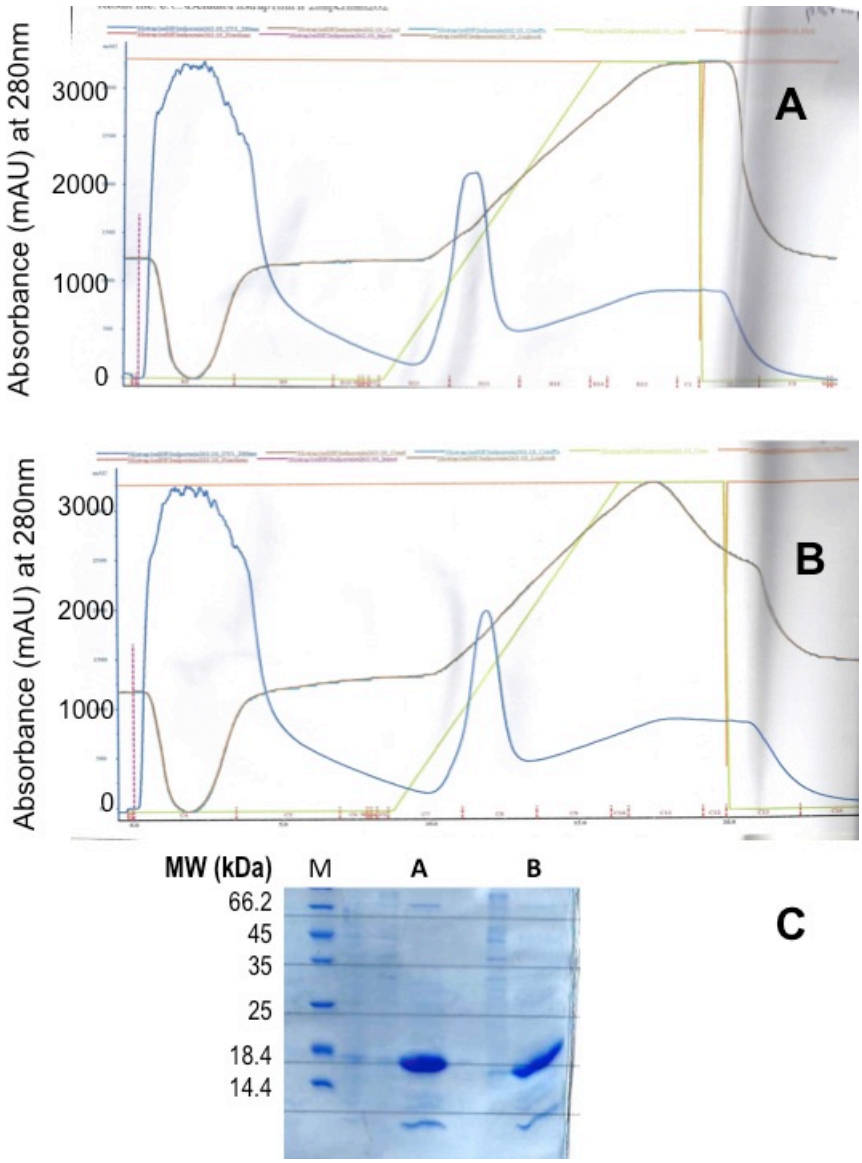


Figure 5.12 (A, B) Elution profiles from a 5 ml His-Trap FF column; the blue line represents A_{280} and the green line indicates the imidazole gradient. (C) Coomassie stained SDS-PAGE analysis of His-tagged human Kibra; 5 µl marker (with band sizes indicated) was loaded in lane M; 10 µl of pooled elution peak fractions diluted 1:1 with 2X sample buffer (Appendix III-c) were run in lanes A and B, which contained ^{15}N-labelled (C_A) WW1-WW2(s) and (C_B) WW1-WW2(l) both of which are around 17 kDa. Lower MW bands are likely to be His-rich contamination from the bacterial cell lysates.

5.3.5 2D NMR conditions and experiments with Kibra WW domains

Tandem WW domains of Kibra exhibit signs of a heterogeneous population in ^{1}H-^{15}N-HSQC spectra with peaks of different intensities and poor dispersion around 9-10 ppm. The spectra also indicate some degree of sample instability/degradation with the appearance of many small peaks. Both short and long constructs interact with single

PPxY peptide. Upon titration with peptide at least four peaks around 9-10 ppm appear in the WW1-WW2(s/l) ^{1}H-^{15}N-HSQC spectra and the intensity of several other peaks increases (Figure 5.13). The evidence indicates that a significant fraction of one WW domain within the protein population undergoes binding-induced folding. The presence of a second Trp residue seven residues C-terminal to the consensus Trp position in WW1-WW2(l) does not appear to change the native structure or affect peptide binding.

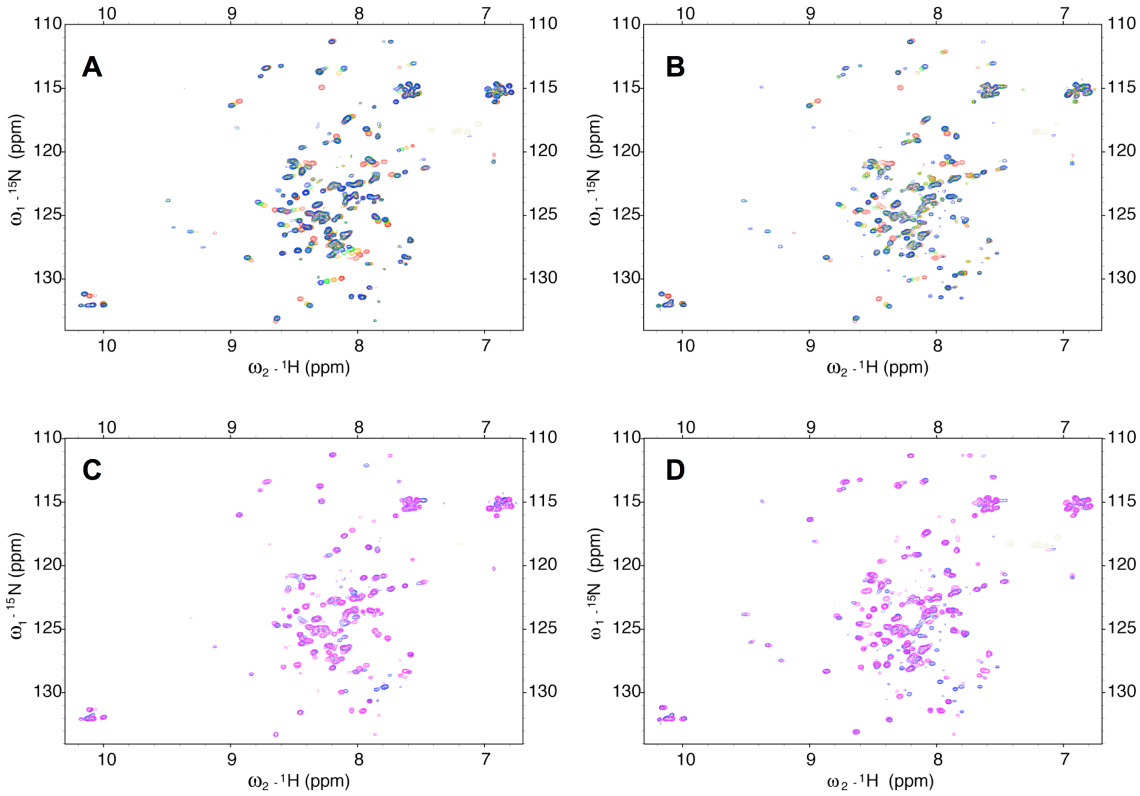


Figure 5.13 (A) Kibra WW1-WW2(s) and **(B)** WW1-WW2(l) titrations with GTPPPPYTVG peptide; red peaks represent the unbound proteins and the final titration point is shown in blue. Also shown is a comparison of WW1-WW2(s) in magenta with WW1-WW2(l) in blue, **(C)** without and **(D)** with PPxY peptide. Native WW1-WW2 appears to be largely unfolded as the peaks are not particularly well dispersed. The peptide induces the appearance of around 6 peaks at 8.5-9.5 ppm which is likely to be representative of only one folded WW domain.

5.4 Discussion

5.4.1 Atypical WW domains of Sav and Kibra

Mouse Sav1 WW2 is unusual as, unlike all other WW domains reported to date, it forms a homodimer (Ohnishi et al. 2007). WW1 from human Magi1, which has a very

similar sequence to Sav1 WW2, does not form a dimer. The homodimerisation of mouse Sav WW2 is thought to be due to the presence of a Ser residue in the first β-strand of the domain, whereas in Magi1 WW1 the Ser is replaced by an Arg residue (Ohnishi et al. 2008). Like mouse Sav1 WW2, *Drosophila* Sav WW2 contains this Ser, however unlike mouse Sav1 WW2, three residues before the Ser a positively charged Arg replaces a negatively charged Glu that is also thought to contribute to the dimerisation. Therefore an attempt to determine whether *Drosophila* Sav WW2 is a monomer or dimer was made.

Glutaraldehyde cross-linking did not indicate that *Drosophila* Sav WW2 forms a dimer. Glutaraldehyde, which is approximately 7.5 Å long, is known to cross-link side chains in Lys-Lys and Lys-Arg pairs (Salem et al. 2010). The structure of *Drosophila* Sav WW2 was modelled upon chains A and B of the mouse Sav1 WW2 dimer (Figure 5.14b,c) using Swiss-Model (Arnold et al. 2006; Kiefer et al. 2009), and PyMol (MacPyMOL 2009-2010) was used to determine the distances between Lys-Arg side chains in the theoretical dimer interface. The shortest distance between Lys-Arg side chains, based on the model, is 7.4 Å (Figure 5.14c). It is possible that the glutaraldehyde could bridge the gap between side chains during dimerisation, so the non-dimerisation observed for *Drosophila* Sav WW2 could be attributed to existing in a monomeric state. However, the structural model is theoretical and if the Lys-Arg side chains were a fraction further apart the glutaraldehyde could not cross link the two molecules of WW2; therefore the possibility that WW2 exists in a dimeric state cannot be ruled out. Ideally mouse Sav1 WW2 would have been used as a positive (dimeric) control, but unfortunately this protein was not available at the time of the experiment.

The replacement of the negatively charged Glu by the positively charged Arg might still be capable of forming a weak salt bridge with the Ser, if this was the only substitution. However, Mouse/human Sav WW2 contains four core aromatic residues, whereas *Drosophila* Sav WW2 only contains two; therefore it is very unlikely that *Drosophila* Sav WW2 forms a dimeric complex.

Mouse Sav1 WW2, which is identical to human Sav1 WW2 in sequence, is also atypical because a Tyr residue replaces the second Trp that is generally conserved within WW domains (and so it could be considered as a 'WY' domain). The replacement of the second Trp is not unique to mouse Sav1 WW2 (Figure 5.3). In *Drosophila* Sav WW2 this residue is replaced by Arg ('WR' domain), in human Kibra WW2 Ile is present ('WI' domain). The NMR titration of *Drosophila* Sav WW2 with

the single generic PPxY peptide (WBP1) revealed no interaction between the atypical WW domain and peptide.

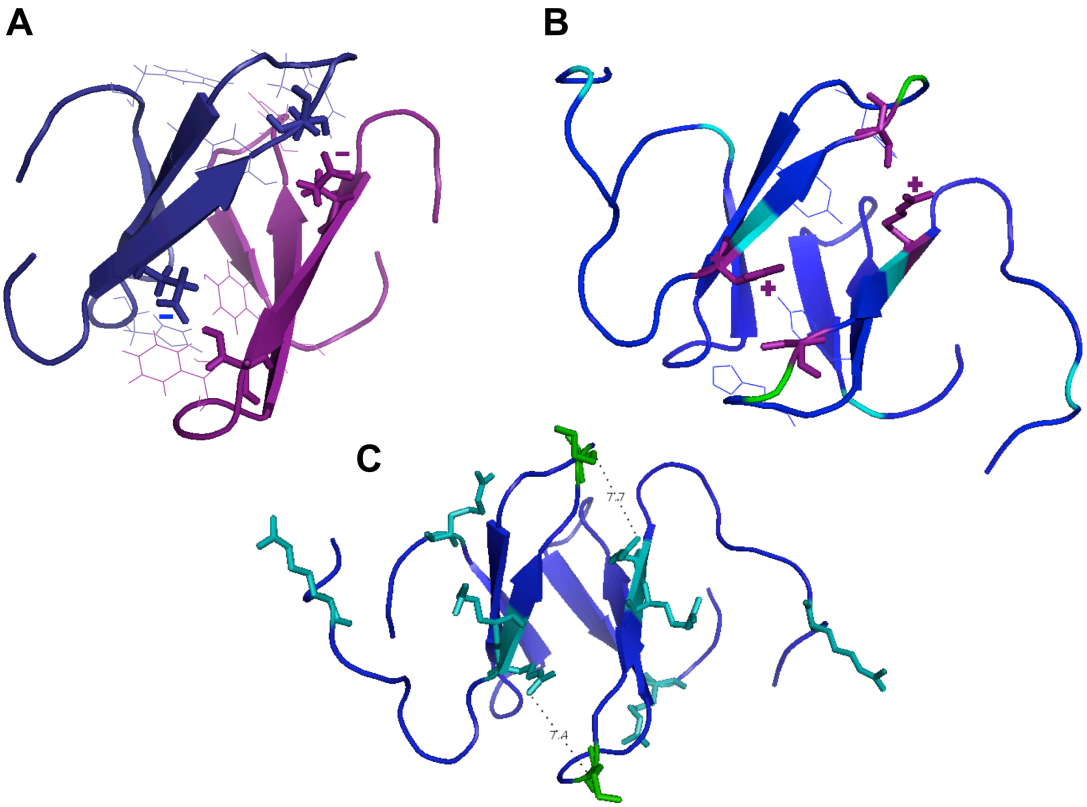


Figure 5.14 (A) Mouse/human Sav WW2 and (B, C) *Drosophila* Sav WW2 models made using Swiss-Model (Arnold et al. 2006; Kiefer et al. 2009) based on the structure of the dimeric mouse Sav1 WW2 (PDB ID: 2DWV (Ohnishi et al. 2007)). (A) Mouse Sav WW2; the polar residues Glu (negative charges shown) and Ser that are located at the N- and C- termini, respectively, of the WW2 β1 strand contribute to the dimerisation and are shown in thick lines. The four aromatic residues (F, Y, H, Y) that also play a role in dimerisation are shown in thin lines. (B) *Drosophila* Sav WW2 dimer model showing Arg (positive charges shown) and Ser residues that are equivalent to the 'dimerisation' residues of mouse Sav WW2. There are fewer aromatic residues in (B) *Drosophila* Sav WW2 compared with (A) mouse Sav WW2. (C) Model of Lys (green) and Arg (cyan) residues that could potentially cross-link with glutaraldehyde and the distance between side chains in the theoretical dimer interface. PyMol (MacPyMOL 2009-2010) was used to make the images.

In the mouse Sav WW2 dimer, the peptide-binding site is obscured. This might prevent protein-protein interactions, or alternatively the as yet unidentified ligand might competitively inhibit homodimerisation. It has frequently been observed that WW domains with replacement residues at the second conserved W are incapable of binding

peptides containing the known WW-binding motifs (Kremerskothen et al. 2003; Fedoroff et al. 2004; Ludes-Meyers et al. 2004). Perhaps because mouse/human Sav WW2 was redundant in terms of normal WW domain functionality, WW1 fulfils this role, it developed the ability to homodimerise to enhance the formation of multi-protein complexes thereby improving the function of Sav as a scaffolding protein.

In addition to the replacement of the conserved Trp by Ile, Kibra WW2 also contains a Trp residue 7 amino acids C-terminal to the WW core. The presence of the C-terminal Trp is not likely to be a splicing/insertion event because the characteristic WW domain conserved Pro is present 3 residues after the Ile (substituted for the second W) and there is not a Pro residue 3 residues after the C-terminal Trp. The presence of the C-terminal Trp does not affect the structure of Kibra WW1-WW2 (Figure 5.13), which might suggest that its proximity to WW2 is coincidental.

As individual Kibra WW domains were not purified it was not possible to investigate the effect of PPxY peptide on the atypical WW2. The titration of Kibra WW1-WW2 with PPxY peptide induced the appearance of four peaks in the 9-10 ppm ^{1}H region, which suggests that only one domain was affected. As it was previously established that WW2 cannot bind a range of peptides with the PPxY motif (Kremerskothen et al. 2003), it is likely that WW1 alone interacted with the peptide and WW2 exists in an unfolded/conformationally dynamic state. Binding motifs for 'Wx' domains (like Kibra and Sav WW2) have been demonstrated for the 'WF' domains of Su(dx) and Smurf2. In the former case, WW4 (WF) of Su(dx) interacts with the AKQPPSYEDC peptide derived from Notch with higher affinity than the other typical WW domains in Su(dx) (45 μM as compared to around 200 μM), but with lower affinity than a 'WF4' to 'WW4' mutant (15 μM) (Jennings et al. 2007). Smurf2 WW3 (WF) was found to bind the ELESPPPPYSRYPMD peptide with a K_d of 40 μM; the six residues C-terminal to the PPxY motif are essential for binding (Chong et al. 2006).

5.4.2 Binding induced folding/stabilisation of Sav and Kibra WW domains

The native *Drosophila* Sav WW1 ^{1}H-^{15}N-HSQC displayed significantly fewer visible NH peaks than expected, probably due to line broadening caused by the domain existing in an equilibrium between folded and partially/completely unfolded states. Upon titration with the single PPxY peptide the *Drosophila* Sav WW1 ^{1}H-^{15}N-HSQC spectrum exhibits well dispersed peaks of similar intensity, characteristic of a folded domain. The binding of peptide by *Drosophila* Sav WW1 probably involves the

population of WW1 molecules that adopt the folded conformation and as the equilibrium shifts, all of the population is brought into the more stable bound folded state. The instability of native *Drosophila* Sav WW1 is further compounded by the presence of lower molecular weight fractions in the purification profile (Figure 5.4a), which are probably caused by degradation or incomplete translation. The latter is less likely as the fragment is very small and a cell strain (Arctic Express (RIL)) was specifically selected to cope with the rare codons.

Binding-induced folding also appears to occur for Kibra WW1-WW2. Equilibrium is also likely to exist between folded and partially folded conformations in this protein. As Kibra WW2 cannot bind a range of PPxY peptides (Kremerskothen et al. 2003) it is highly likely that interactions between PPxY peptide and WW1 are responsible for the extra peaks observed in the ^{1}H-^{15}N-HSQC. As WW1 is almost certainly the domain that undergoes binding induced folding, the lack of well dispersed peaks in the native Kibra WW1-WW2 spectra suggest that WW2 is probably natively unfolded or in equilibrium between conformations. Previously the consensus sequence observed to have the greatest affinity with Kibra WW1 was RxPPxY, however our peptide of TxPPxY was evidently capable of interacting with the WW domain.

Another case where titration of PPxY peptide was necessary for complete WW domain folding was that of Su(dx) (Fedoroff et al. 2004). Native Su(dx) WW3-WW4 does not adopt a stable folded conformation, WW4 is a 'WF' domain and does not bind peptide. WW3-PPxY binding stabilised the whole tandem domain structure thus allowing for NH backbone assignments and subsequent structure calculations. Only 5 weak peaks appeared around 9-10 ppm during the Kibra WW1-WW2 titration, indicating the presence of only one fully folded WW domain. As only one WW domain is stabilised by peptide, Kibra does not appear to behave in a similar manner to Su(dx), it is more like *Drosophila* Sav with independent functional WW1 domains and WW2 domains. The function of the atypical Kibra WW2 is undetermined. Kibra WW1-WW2, like Sav WW1-WW2, exhibits signs of degradation indicated by the presence of many small peaks in the ^{1}H-^{15}N-HSQC spectra. Proteolysis is more prevalent in unfolded proteins, so the possible conformational instability in Sav and Kibra WW1-WW2 could be promoting degradation of the recombinant proteins.

CHAPTER 6: FINAL DISCUSSION AND CONCLUSIONS

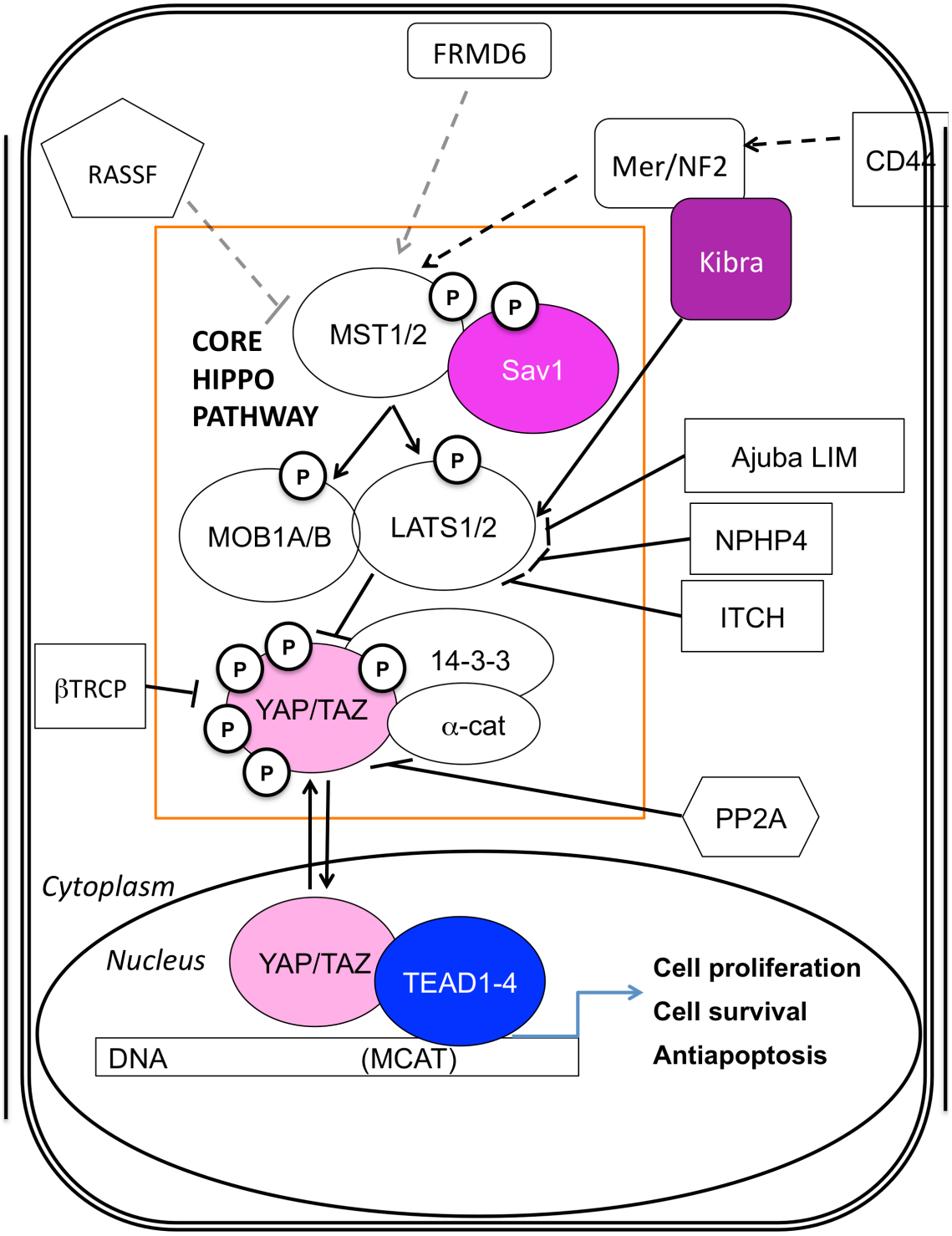


Figure 6.1 Location within the Hippo pathway of all proteins studied i.e. TEAD (blue), YAP/TAZ (pink), Sav1 (magenta), and Kibra (purple). Core Hippo pathway components are located within the orange box.

6.1 Discussion

Nuclear targets of the Hippo pathway: TEAD-YAP/TAZ. The TEAD transcription factors form an active complex with the Hippo pathway target proteins YAP and TAZ to mediate transcription of genes that promote cell proliferation, survival and differentiation. Our attempts to characterise the two functional domains of TEAD were ultimately unsuccessful, though given more time success might have been achieved. The crystal structures of YBDs from several TEAD proteins both alone and in complex with fragments of YAP are now known. The structures will enable more detailed analysis of the TEAD family of transcription factors and their binding partners.

Alternative isoforms and repeat regions. In the attempt to generate cDNA for subsequent cloning of medaka fish TEAD protein domains, alternate isoforms of medaka TEAD1, 2/3 and 4 were identified. The TEAD4 variant has a stop codon between the DBD and the YBD, potentially leading to transcription of a truncated non-YAP-binding protein. TEAD1 and TEAD 2/3 variants contain a 20 residue insertion C-terminal to the DBD. This insertion is highly similar to the H3 helix that contains the NLS and mediates DNA binding. Due to this similarity the insertion is probably α-helical in structure. Cooperativity between multiple NLSs to enhance transportation into the nucleus has been observed in the human mismatch repair protein MSH6 (Gassman et al. 2011). It is conceivable that the role of the insertion in these TEAD variants is to enhance nuclear localisation, particularly as the transcription factors are also responsible for the nuclear transport of their co-activators YAP and TAZ. With the extra region being in such close proximity to the conserved DBD, it would be interesting to assess the effect on TEAD-DNA interactions.

Regional repeats appear frequently in proteins of the Hippo pathway. In addition to the identification of TEAD variants during cDNA generation, a second TAZ isoform was also found. Like its paralogue YAP, the second splice form of TAZ contains a second WW domain. One difference between the tandem domains of YAP and TAZ is that the newly discovered TAZ2 has a longer linker by around 10 residues. When repeat regions have a known function, the question of potential cooperativity often arises (as was speculated on for the TEAD insertion). Often the distance between domains and the structural features of the inter-domain linker has a marked effect on such cooperativity.

NMR spectroscopy was used to investigate structural features of the YAP and TAZ tandem WW domains and potential inter-WW domain cooperation.

NMR studies of tandem WW domains. The inter-WW domain linker was found to be unstructured and flexible in both YAP and TAZ. Perhaps unsurprisingly there was more freedom of movement in TAZ, with the longer linker, than YAP. Other NMR-based studies of tandem WW domains with unstructured linkers found more restriction of internal motion than we observed, however the linkers of YAP and TAZ are significantly longer than in any of the other characterised proteins so again this observation is not particularly unexpected.

A recently published solution structure of YAP WW1-WW2 in complex with a phospho-Ser-PPxY peptide confirmed our finding that the inter-domain linker was a flexible loop, however the structure also contained an α-helix at either end of the linker (Aragon et al. 2011). The $\{^{1}H\}$-^{15}N-NOE values for YAP and TAZ observed in Chapter 4 reached the minimum level (close to or below 0) approximately 10 residues after the core of the first WW domain and began increasing approximately 10 residues before the second WW. This gradual decrease/increase in NOE values could be consistent with restricted movement imposed by helices. The structure of medaka YAP WW2 appeared to contain a helical coil N-terminal to the core WW domain, though the orientation was not superimposed on all 20 of the models. This may also be an indication the C-terminus of the inter WW-linker contains some residual structure. Individual YAP WW1 domains in the PDB do not contain enough linker residues to determine if a similar effect is observed C-terminal to the WW1 core (i.e. the beginning of the linker).

The dual PPxY peptide did not affect the inter-WW linker meaning potential helices connecting the WW domains to the linker are likely to also be present in the native protein. The dual PPxY peptide did not induce the formation of extra structural changes in comparison to the single PPxY peptide, indicating that the linker plays no active role in peptide binding. The possibility that other dual WW-binding site peptides could induce allosteric conformational changes in the tandem WW domains of YAP and TAZ cannot be excluded. Although the linker region of YAP and TAZ WW1-WW2 plays no active role in peptide binding, at least six-fold increased affinity for dual PPxY peptide was observed over single PPxY peptide. This is a smaller effect than the thirty-fold increase in affinity observed for FBP21 binding to dual Pro-rich peptide (Huang et al. 2009). FBP21 contains a shorter linker (12 residues) than YAP and TAZ; we did

observe a slightly higher affinity in YAP than TAZ for single PPxY peptide, which may be an effect caused by the shorter YAP linker.

Enhanced and cooperative binding through tandem WW domains. Su(dx) WW3 and WW4 exhibit cooperativity whereby WW3-WW4 affinity for a WW3-binding peptide is higher than WW3 alone (Jennings et al. 2007). Binding of peptide to Su(dx) WW3 (within WW3-WW4) leads to folding of WW4, which subsequently increases the affinity of WW4 for its peptide. The cooperative effects of ligand binding to Su(dx) are due to interactions between the native unbound WW3 and WW4 (Jennings et al. 2007). Evidence indicated that YAP/TAZ WW1 and WW2 do not interact and Aragon et al (2011) also reported a lack of YAP inter-WW contacts. Cooperativity is not likely to contribute to enhanced binding in the dual PPxY peptide as chemical shift changes occurred concurrently in the two WW domains, rather than sequentially. One explanation for the observed affinity effect is an avidity-based mechanism, whereby binding to one WW domain brings the second PPxY into close proximity with the second WW domain. This theory would also apply to the increased affinity observed in the case of FPB21 (Huang et al. 2009).

Atypical WW domains. The upstream Hippo pathway proteins Sav and Kibra contain tandem WW domains without linker regions. The C-terminal WW domains of both Sav and Kibra are atypical 'Wx' domains, like Su(dx) WW4. Unfortunately canonical binding motifs for the 'Wx' domains of Sav and Kibra WW2 have not been identified, and in the case of the potentially dimeric mouse Sav WW2 might not exist. *Drosophila* Sav WW2 contains several differences in key 'dimerisation' residues compared with mouse Sav, and based upon these differences *Drosophila* Sav WW2 is predicted to be monomeric, though this has not been definitively proven.

Due to the lack of known 'Wx' binding motifs it is not possible to assess the structural impact of peptide binding using experiments such as those used for Su(dx) to assess the potential effects of cooperativity between two different ligands and the two WW domains. A second limitation of studying these tandem WW domains is sample instability. The poor ^{1}H-^{15}N-HSQC spectra seem to indicate that the atypical Kibra WW2 within WW1-WW2 is unfolded and the major population of *Drosophila* Sav WW2 domain within WW1-WW2 is natively folded. In both Sav and Kibra the typical group I N-terminal WW domains may be partially unfolded and undergo conformational exchange. The instability of the Sav and Kibra WW1-WW2 samples

makes it difficult to draw any definite conclusions about the conformation of the atypical tandem WW domains. Mixing experiments with the individual ^{15}N labelled Sav WW2 and unlabelled Sav WW1 revealed no chemical shift changes, indicating that the two domains do not interact.

Conformational flexibility. Besides the general instability of the tandem WW domain samples (Sav and Kibra in particular), most of the Hippo pathway WW domains studied exhibited some degree of conformational exchange. YAP (WW1), TAZ (WW1), Sav (WW1, WW1-WW2, WW2) and Kibra (WW1-WW2) all display weak peaks around proton chemical shift values of 9-10 ppm in ^{1}H-^{15}N-HSQC spectra; the difference in intensity to other peaks suggests that the whole population is not in one fully folded conformation. In the case of YAP WW1, ^{15}N R_2 transverse relaxation data and RMSD within R_2 dispersion data indicated that the domain undergoes significant conformational exchange with residues facing the stabilizing 'buckle' opposite the peptide-binding face being affected. The TAZ WW1-WW2 ^{1}H-^{15}N-HSQC spectrum displayed broad line widths and relaxation experiments (^{15}N R_1 and ^{15}N R_2) showed that both WW domains are subject to conformational instability.

The low number of NH peaks in the *Drosophila* Sav WW1 ^{1}H-^{15}N-HSQC spectrum suggests that the domain is in equilibrium between unfolded/partially folded and fully folded conformations. Presumably, as the structure of mouse Sav WW1 (which has the same sequence as human Sav WW1) is present in the PDB it is more stable than its *Drosophila* counterpart. The Kibra WW1-WW2 ^{1}H-^{15}N-HSQC spectrum also displays signs of heterogeneity with variable peak intensities. The ^{1}H-^{15}N-HSQC spectra from Sav WW1, Kibra WW1-WW2, YAP WW1-WW2, and TAZ WW1-WW2 all improve by varying degrees upon the addition of PPxY peptide. The most obvious improvement was for Sav WW1 where a few peaks in the centre of the spectrum were transformed upon the addition of peptide into a well dispersed pattern consistent with a homogeneous folded domain.

It would appear that the WW domains undergoing conformational exchange are subject to a certain degree of binding-induced folding. The most likely scenario is conformational selection where peptide binds to the folded fraction, shifting the equilibrium until the whole population is folded. Binding-induced folding of WW domains is likely to have functional implications for Hippo pathway proteins. It is possible that the conformational instability of the WW domains contributes to binding specificity. For example YAP WW1, which is dynamic in terms of conformation, binds

the phospho-Ser site of the Smad1 peptide (Figure 4.17) and WW2, which exists in a single conformation, preferentially binds to the PPxY site on the same peptide (Aragon et al. 2011). WW domain interactions of YAP, TAZ, Sav and Kibra mediate cross-talk between signalling pathways; conformational exchange to alter binding partner specificity may be one mechanism that enables switching between pathways.

The mechanisms of WW domain folding and general stability are perhaps not yet completely predictable despite a large number of experimental and theoretical studies. One molecular dynamics simulation found that the stability of WW domains is affected by various factors including electrostatic interactions, side-chain packing, conformational entropy of the flexible regions either end of the WW core, and the binding of cognate ligand (Ibragimova et al. 1999). Given the small size of these domains it is not surprising that minor differences can significantly affect the structural stability and binding preferences. However, the prevalence of WW domains in intracellular proteins involved in transcriptional regulation highlights them as essential protein modules worthy of detailed investigation.

6.2 Conclusions

6.2.1 TEAD domains and TBDs of YAP and TAZ.

Variant isoforms of medaka fish TEAD mRNA were identified, though the biological relevance of these isoforms remains undetermined. TEAD domains (DBD and YBD) are intrinsically structured, though mostly insoluble when expressed in *E. coli* and therefore difficult to work with. TBDs of YAP and TAZ are natively unstructured and are also insoluble in the absence of the C-terminal WW domain.

6.2.2 YAP and TAZ WW domains

A second isoform of medaka TAZ was discovered, TAZ2, which contains a second WW domain much like YAP2. The medaka YAP WW2 structure is similar to other WW domains. WW1 of YAP and both WW domains of TAZ undergo conformational exchange but appear to retain the classical WW domain structure in the predominant folded states. YAP and TAZ WW domains, within WW1-WW2 pairs, are capable of binding PPxY peptide with similar affinity. The presence of a second PPxY motif

within a titrated peptide increases the affinity of the tandem WW domain pairs for ligand by at least 6-times. The WW domain linker in TAZ2 is longer than in YAP2 by 10 amino acids. The linker region is highly flexible in the two proteins, in both free and ligand bound states.

6.2.3 Sav and Kibra WW domains

Drosophila Sav WW2 is stable, natively folded, and cannot bind PPxY peptide. *Drosophila* Sav WW1 appears to be in equilibrium between folded and unfolded states but the domain retains its stable secondary structure when bound to PPxY peptide. The tandem WW domain constructs of *Drosophila* and human Sav and human Kibra are unstable, heterogeneous, and partially unfolded in their native state. *Drosophila* Sav WW1-WW2 undergoes partial folding (which involves WW1 only) upon titration with PPxY peptide but heterogeneity within the NMR sample remains. WW1 within Kibra WW1-WW2 seems to gain more stability by binding induced folding during ligand titration, however the samples do not appear to become completely homogeneous and WW2 does not appear to be folded.

6.3 Future directions

Our intended short-term outcomes of the TEAD-YAP studies have been thoroughly covered by three publications as well as relevant reviews. Although it would be interesting to further study the intricacies of the various TEAD proteins, perhaps by modelling them using the available structures, and studying how interactions with YAP and TAZ differ, it would not be efficient to pursue this avenue of investigation at the current time. As for long-term goals, an ambitious but worthwhile task would be to study ternary DNA-TEAD-YAP and DNA-TEAD-TAZ complexes and elucidate the exact role of the C-terminal transactivation domain from YAP/TAZ. As both TEAD and YAP/TAZ are required for transcriptional activation, it would also be interesting to determine whether cooperativity between the DBD and YBD of TEAD is involved in binding YAP/TAZ and DNA. The WW domains of YAP/TAZ are known to be promiscuous; another area of investigation could be to determine whether they interact with transcriptional machinery to form an intra-nuclear multi-protein complex.

Our publication of the YAP and TAZ WW domain data shows that this study reached a satisfactory outcome, though it has raised more questions about the conformational stability of WW domains and their numerous roles within various cell-signalling networks. Several studies have highlighted the importance of residues flanking the PPxY motif with regard to affinity; it would therefore be interesting to assess the effect of varying the peptide sequence on the YAP/TAZ WW domain pair. Altering the linker length between PPxY motifs and WW domains would also increase understanding of the mechanisms of enhanced binding via multiple WW domain activity.

There are many questions to be answered about the content of Chapter 4, but there are also many difficulties to overcome with the solubility and general stability of the Sav and Kibra WW domains. In order to make the Sav and Kibra work truly comparable in terms of Hippo pathway function, the constructs should be derived from the same organism. Therefore, it would be useful to clone, express and purify the individual WW domains of human Sav to assess whether they behave in the same manner as their *Drosophila* counterparts. It would also be useful to produce individual WW domain constructs for Kibra and to optimise buffer systems for all of the protein fragments for further characterisation by NMR. A long-term goal would be to identify binding motifs for the atypical 'Wx' domains of Sav and Kibra, compare the interaction of such partners with typical WW domains, and further explore the possibilities of inter-WW domain cooperation.

Clearly the proteins of the Hippo pathway contain multiple protein-protein interaction domains that could potentially mediate the formation of huge multi-protein complexes. We have used the most practical approach of 'divide and conquer' with regard to characterising the various interaction modules. Ultimately the goal would be to build up a larger picture of the proteins involved in the formation of such complexes at any given time and assess the interactions that regulate them to gain a systems level understanding of the processes involved in organ growth, tumour suppression and other physiological processes that the Hippo pathway impacts upon.

CHAPTER 7: REFERENCES

Alarcon, C., A. I. Zaromytidou, Q. R. Xi, S. Gao, J. Z. Yu, S. Fujisawa, . . . J. Massague (2009). "Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF-beta Pathways." Cell **139**(4): 757-769.

Anbanandam, A., D. C. Albarado, C. T. Nguyen, G. Halder, X. L. Gao and S. Veeraraghavan (2006). "Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain." Proceedings of the National Academy of Sciences of the United States of America **103**(46): 17225-17230.

Aragon, E., N. Goerner, A. I. Zaromytidou, Q. Xi, A. Escobedo, J. Massague and M. J. Macias (2011). "A Smad action turnover switch operated by WW domain readers of a phosphoserine code." Genes and Development **25**(12): 1275-1288.

Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.

Avruch, J., D. Zhou, J. Fitamant and N. Bardeesy (2011). "Mst1/2 signalling to Yap: gatekeeper for liver size and tumour development." British Journal of Cancer **104**(1): 24-32.

Badouel, C., L. Gardano, N. Amin, A. Garg, R. Rosenfeld, T. Le Bihan and H. McNeill (2009). "The FERM-domain protein Expanded regulates Hippo pathway activity via direct interactions with the transcriptional activator Yorkie." Developmental Cell **16**(3): 411-420.

Badouel, C., A. Garg and H. McNeill (2009). "Herding Hippos: regulating growth in flies and man." Current Opinion in Cell Biology **21**(6): 837-843.

Ball, L. J., R. Kuhne, J. Schneider-Mergener and H. Oschkinat (2005). "Recognition of proline-rich motifs by protein-protein-interaction domains." Angewandte Chemie-International Edition **44**(19): 2852-2869.

Bao, Y., Y. Hata, M. Ikeda and K. Withanage (2011). "Mammalian Hippo pathway: from development to cancer and beyond." Journal of Biochemistry **149**(4): 361-379.

Basu, S., N. F. Totty, M. S. Irwin, M. Sudol and J. Downward (2003). "Akt phosphorylates the Yes-associated protein, YAP, to induce interaction with 14-3-3 and attenuation of p73-mediated apoptosis." Molecular Cell **11**(1): 11-23.

Chan, S. W., C. J. Lim, Y. F. Chong, A. V. Pobbati, C. Huang and W. Hong (2011). "Hippo pathway-independent restriction of TAZ and YAP by angiomotin." Journal of Biological Chemistry **286**(9): 7018-7026.

Chan, S. W., C. J. Lim, L. S. Loo, Y. F. Chong, C. Huang and W. Hong (2009). "TEADs mediate nuclear retention of TAZ to promote oncogenic transformation." Journal of Biological Chemistry **284**(21): 14347-14358.

Chen, H. I. and M. Sudol (1995). "The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules." Proceedings of the National Academy of Sciences of the United States of America **92**(17): 7819-7823.

Chen, L., P. G. Loh and H. Song (2010). "Structural and functional insights into the TEAD-YAP complex in the Hippo signaling pathway." Protein Cell **1**(12): 1073-1083.

Chen, L. M., S. W. Chan, X. Q. Zhang, M. Walsh, C. J. Lim, W. J. Hong and H. W. Song (2010). "Structural basis of YAP recognition by TEAD4 in the Hippo pathway." Genes & Development **24**(3): 290-300.

Choi, J., S. Oh, D. Lee, H. J. Oh, J. Y. Park, S. B. Lee and D. S. Lim (2009). "Mst1-FoxO signaling protects Naive T lymphocytes from cellular oxidative stress in mice." PLoS One **4**(11): e8011.

Chong, P. A., H. Lin, J. L. Wrana and J. D. Forman-Kay (2006). "An expanded WW domain recognition motif revealed by the interaction between Smad7 and the E3 ubiquitin ligase Smurf2." Journal of Biological Chemistry **281**(25): 17069-17075.

Chong, P. A., H. Lin, J. L. Wrana and J. D. Forman-Kay (2010). "Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity." Proceedings of the National Academy of Sciences of the United States of America **107**(43): 18404-18409.

Cole, C., J. D. Barber and G. J. Barton (2008). "The Jpred 3 secondary structure prediction server." Nucleic Acids Research **36**(Web Server issue): W197-201.

Das Thakur, M., Y. Feng, R. Jagannathan, M. J. Seppa, J. B. Skeath and G. D. Longmore (2010). "Ajuba LIM proteins are negative regulators of the Hippo signaling pathway." Current Biology **20**(7): 657-662.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." Journal of Biomolecular NMR **6**(3): 277-293.

Dong, J. X., G. Feldmann, J. B. Huang, S. Wu, N. L. Zhang, S. A. Comerford, . . . D. J. Pan (2007). "Elucidation of a universal size-control mechanism in Drosophila and mammals." Cell **130**: 1120-1133.

Dupont, S., L. Morsut, M. Aragona, E. Enzo, S. Giulitti, M. Cordenonsi, . . . S. Piccolo (2011). "Role of YAP/TAZ in mechanotransduction." Nature **474**(7350): 179-183.

Ellenberger, T., D. Fass, M. Arnaud and S. C. Harrison (1994). "Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer." Genes and Development **8**(8): 970-980.

Espanel, X. and M. Sudol (1999). "A single point mutation in a Group IWW domain shifts its specificity to that of Group IIWW domains." Journal of Biological Chemistry **274**(24): 17284-17289.

Espanel, X. and M. Sudol (2001). "Yes-associated protein and p53-binding protein-2 interact through their WW and SH3 domains." Journal of Biological Chemistry **276**(17): 14514-14523.

Fedoroff, O. Y., S. A. Townson, A. P. Golovanov, M. Baron and J. M. Avis (2004). "The structure and dynamics of tandem WW domains in a negative regulator of notch signaling, suppressor of deltex." Journal of Biological Chemistry **279**(33): 34991-35000.

Fernandez, L. A., P. A. Northcott, J. Dalton, C. Fraga, D. Ellison, S. Angers, . . . A. M. Kenney (2009). "YAP1 is amplified and up-regulated in hedgehog-associated medulloblastomas and mediates Sonic hedgehog-driven neural precursor proliferation." Genes and Development **23**(23): 2729-2741.

Ferre-D'Amare, A. R., P. Pognonec, R. G. Roeder and S. K. Burley (1994). "Structure and function of the b/HLH/Z domain of USF." EMBO Journal **13**(1): 180-189.

Ferrigno, O., F. Lallemand, F. Verrecchia, S. L'Hoste, J. Camonis, A. Atfi and A. Mauviel (2002). "Yes-associated protein (YAP65) interacts with Smad7 and potentiates its inhibitory activity against TGF-beta/Smad signaling." Oncogene **21**(32): 4879-4884.

Fielding, L. (2007). "NMR methods for the determination of protein-ligand dissociation constants." Progress in Nuclear Magnetic Resonance Spectroscopy **51**(4): 219-242.

Gassman, N. R., J. E. Clodfelter, A. K. McCauley, K. Bonin, F. R. Salsbury, Jr. and K. D. Scarpinato (2011). "Cooperative nuclear localization sequences lend a novel role to the N-terminal region of MSH6." PLoS One **6**(3): e17907.

Gavva, N. R., R. Gavva, K. Ermekova, M. Sudol and C. K. J. Shen (1997). "Interaction of WW domains with hematopoietic transcription factor p45/NF-E2 and RNA polymerase II." Journal of Biological Chemistry **272**(39): 24105-24108.

Genevet, A., M. C. Wehr, R. Brain, B. J. Thompson and N. Tapon (2010). "Kibra Is a Regulator of the Salvador/Warts/Hippo Signaling Network." Developmental Cell **18**(2): 300-308.

Goddard, T. D. and D. G. Kneller SPARKY 3. San Francisco, University of California.

Goulev, Y., J. D. Fauny, B. Gonzalez-Marti, D. Flagiello, J. Silber and A. Zider (2008). "Scalloped interacts with Yorkie, the nuclear effector of the Hippo tumor-suppressor pathway in *Drosophila*." Current Biology **18**(6): 435-441.

Habbig, S., M. P. Bartram, R. U. Muller, R. Schwarz, N. Andriopoulos, S. Chen, . . . B. Schermer (2011). "NPHP4, a cilia-associated protein, negatively regulates the Hippo pathway." Journal of Cell Biology **193**(4): 633-642.

Halder, G. and R. L. Johnson (2011). "Hippo signaling: growth control and beyond." Development **138**(1): 9-22.

Hao, Y., A. Chun, K. Cheung, B. Rashidi and X. Yang (2008). "Tumor suppressor LATS1 is a negative regulator of oncogene YAP." Journal of Biological Chemistry **283**(9): 5496-5509.

Hilman, D. and U. Gat (2011). "The Evolutionary History of YAP and the Hippo/YAP Pathway." Molecular Biology and Evolution **28**(8): 2403-2417.

Hong, J. H., E. S. Hwang, M. T. McManus, A. Amsterdam, Y. Tian, R. Kalmukova, . . . M. B. Yaffe (2005). "TAZ, a transcriptional modulator of mesenchymal stem cell differentiation." Science **309**(5737): 1074-1078.

Howell, M., C. Borchers and S. L. Milgram (2004). "Heterogeneous nuclear ribonuclear protein U associates with YAP and regulates its co-activation of Bax transcription." Journal of Biological Chemistry **279**(25): 26300-26306.

Huang, X. J., M. Beullens, J. H. Zhang, Y. Zhou, E. Nicolaescu, B. Lesage, . . . Y. Y. Shi (2009). "Structure and Function of the Two Tandem WW Domains of the Pre-mRNA Splicing Factor FBP21 (Formin-binding Protein 21)." Journal of Biological Chemistry **284**(37): 25375-25387.

Ibragimova, G. T. and R. C. Wade (1999). "Stability of the beta-sheet of the WW domain: A molecular dynamics simulation study." Biophysical Journal **77**(4): 2191-2198.

Ilsley, J. L., M. Sudol and S. J. Winder (2002). "The WW domain: linking cell signalling to the membrane cytoskeleton." Cellular Signalling **14**(3): 183-189.

Jennings, M. D., R. T. Blankley, M. Baron, A. P. Golovanov and J. M. Avis (2007). "Specificity and autoregulation of notch binding by tandem WW domains in Suppressor of Deltex." Journal of Biological Chemistry **282**: 29032-29042.

Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.

Kanai, F., P. A. Marignani, D. Sarbassova, R. Yagi, R. A. Hall, M. Donowitz, . . . M. B. Yaffe (2000). "TAZ: a novel transcriptional co-activator regulated by interactions with 14-3-3 and PDZ domain proteins." Embo Journal **19**(24): 6778-6791.

Kaneko, K. J., E. B. Cullinan, K. E. Latham and M. L. DePamphilis (1997). "Transcription factor mTEAD-2 is selectively expressed at the beginning of zygotic gene expression in the mouse." Development **124**(10): 1963-1973.

Kango-Singh, M. and A. Singh (2009). "Regulation of organ size: insights from the Drosophila Hippo signaling pathway." Developmental Dynamics **238**(7): 1627-1637.

Karpowicz, P., J. Perez and N. Perrimon (2010). "The Hippo tumor suppressor pathway regulates intestinal stem cell regeneration." Development **137**(24): 4135-4145.

Kiefer, F., K. Arnold, M. Kunzli, L. Bordoli and T. Schwede (2009). "The SWISS-MODEL Repository and associated resources." Nucleic Acids Research **37**(Database issue): D387-392.

Kilili, G. K. and J. M. Kyriakis (2010). "Mammalian Ste20-like kinase (Mst2) indirectly supports Raf-1/ERK pathway activity via maintenance of protein phosphatase-2A catalytic subunit levels and consequent suppression of inhibitory Raf-1 phosphorylation." Journal of Biological Chemistry **285**(20): 15076-15087.

Kitagawa, M. (2007). "A Sveinsson's chorioretinal atrophy-associated missense mutation in mouse Tead1 affects its interaction with the co-factors YAP and TAZ." Biochemical and Biophysical Research Communications **361**(4): 1022-1026.

Koepf, E. K., H. M. Petrassi, G. Ratnaswamy, M. E. Huff, M. Sudol and J. W. Kelly (1999). "Characterization of the structure and function of W -> FWW domain variants: Identification of a natively unfolded protein that folds upon ligand binding." Biochemistry **38**(43): 14338-14351.

Komuro, A., M. Nagai, N. E. Navin and M. Sudol (2003). "WW domain-containing protein YAP associates with ErbB-4 and acts as a co-transcriptional activator for the carboxyl-terminal fragment of ErbB-4 that translocates to the nucleus." Journal of Biological Chemistry **278**(35): 33334-33341.

Kremerskothen, J., C. Plaas, K. Buther, I. Finger, S. Veltel, T. Matanis, . . . A. Barnekow (2003). "Characterization of KIBRA, a novel WW domain-containing protein." Biochemical and Biophysical Research Communications **300**(4): 862-867.

Kulman, J. D., J. E. Harris, L. Xie and E. W. Davie (2007). "Proline-rich Gla protein 2 is a cell-surface vitamin K-dependent protein that binds to the transcriptional coactivator Yes-associated protein." Proceedings of the National Academy of Sciences of the United States of America **104**(21): 8767-8772.

Laskowski, R. A., J. A. C. Rullmann, M. W. MacArthur, R. Kaptein and J. M. Thornton (1996). "AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR." Journal of Biomolecular NMR **8**(4): 477-486.

Laughon, A. (1991). "DNA binding specificity of homeodomains." Biochemistry **30**(48): 11357-11367.

Lei, Q. Y., H. Zhang, B. Zhao, Z. Y. Zha, F. Bai, X. H. Pei, . . . K. L. Guan (2008). "TAZ promotes cell proliferation and epithelial-mesenchymal transition and is inhibited by the Hippo pathway." Molecular and Cellular Biology **28**(7): 2426-2436.

Lepre, C. A. and J. M. Moore (1998). "Microdrop screening: A rapid method to optimize solvent conditions for NMR spectroscopy of proteins." Journal of Biomolecular Nmr **12**(4): 493-499.

Li, Z., B. Zhao, P. Wang, F. Chen, Z. H. Dong, H. R. Yang, . . . Y. H. Xu (2010). "Structural insights into the YAP and TEAD complex." Genes & Development **24**(3): 235-240.

Liu, C. Y., X. Lv, T. Li, Y. Xu, X. Zhou, S. Zhao, . . . K. L. Guan (2011). "PP1 cooperates with ASPP2 to dephosphorylate and activate TAZ." Journal of Biological Chemistry **286**(7): 5558-5566.

Long, F., A. A. Vagin, P. Young and G. N. Murshudov (2008). "BALBES: a molecular-replacement pipeline." Acta Crystallographica Section D: Biological Crystallography **64**(Pt 1): 125-132.

Ludes-Meyers, J. H., H. Kil, A. K. Bednarek, J. Drake, M. T. Bedford and C. M. Aldaz (2004). "WWOX binds the specific proline-rich ligand PPXY: identification of candidate interacting proteins." Oncogene **23**(29): 5049-5055.

Luo, X. (2010). "Snapshots of a hybrid transcription factor in the Hippo pathway." Protein Cell **1**(9): 811-819.

Macias, M. J., M. Hyvonen, E. Baraldi, J. Schultz, M. Sudol, M. Saraste and H. Oschkinat (1996). "Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide." Nature **382**(6592): 646-649.

MacPyMOL (2009-2010). The PyMOL Molecular Graphics System, Schrödinger, LLC.

Mohler, P. J., S. M. Kreda, R. C. Boucher, M. Sudol, M. J. Stutts and S. L. Milgram (1999). "Yes-associated protein 65 localizes p62(c-Yes) to the apical compartment of airway epithelia by association with EBP50." Journal of Cell Biology **147**(4): 879-890.

Nicolay, B. N., B. Bayarmagnai, A. B. Islam, N. Lopez-Bigas and M. V. Frolov (2011). "Cooperation between dE2F1 and Yki/Sd defines a distinct transcriptional program necessary to bypass cell cycle exit." Genes and Development **25**(4): 323-335.

Ohnishi, S., P. Guntert, S. Koshiba, T. Tomizawa, R. Akasaka, N. Tochio, . . . S. Yokoyama (2007). "Solution structure of an atypical WW domain in a novel beta-clam-like dimeric form." FEBS Letters **581**(3): 462-468.

Ohnishi, S., N. Tochio, T. Tomizawa, R. Akasaka, T. Harada, E. Seki, . . . S. Yokoyama (2008). "Structural basis for controlling the dimerization and stability of the WW domains of an atypical subfamily." Protein Science **17**(9): 1531-1541.

Oka, T., V. Mazack and M. Sudol (2008). "Mst2 and Lats kinases regulate apoptotic function of YAP." Journal of Biological Chemistry **283**(41): 27534-27546.

Oka, T., E. Remue, K. Meerschaert, B. Vanloo, C. Boucherie, D. Gfeller, . . . M. Sudol (2010). "Functional complexes between YAP2 and ZO-2 are PDZ domain-dependent, and regulate YAP2 nuclear localization and signalling." Biochemical Journal **432**(3): 461-472.

Oka, T. and M. Sudol (2009). "Nuclear localization and pro-apoptotic signaling of YAP2 require intact PDZ-binding motif." Genes to Cells **14**(5): 607-615.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods in Enzymology **276**: 307-326.

Parsons, L. M., N. A. Grzeschik, M. L. Allott and H. E. Richardson (2010). "Lgl/aPKC and Crb regulate the Salvador/Warts/Hippo pathway." Fly (Austin) **4**(4): 288-293.

Pires, J. R., F. Taha-Nejad, F. Toepert, T. Ast, U. Hoffmuller, J. Schneider-Mergener, . . . L. Oschkinat (2001). "Solution structures of the YAP65 WW domain and the variant L30 K in complex with the peptides GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope." Journal of Molecular Biology **314**(5): 1147-1156.

Piserchio, A., R. Ghose and D. Cowburn (2009). "Optimized bacterial expression and purification of the c-Src catalytic domain for solution NMR studies." Journal of Biomolecular NMR **44**(2): 87-93.

Potterton, E., P. Briggs, M. Turkenburg and E. Dodson (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallographica Section D: Biological Crystallography **59**(Pt 7): 1131-1137.

Reddy, B. V., C. Rauskolb and K. D. Irvine (2010). "Influence of fat-hippo and notch signaling on the proliferation and differentiation of Drosophila optic neuroepithelia." Development **137**(14): 2397-2408.

Remue, E., K. Meerschaert, T. Oka, C. Boucherie, J. Vandekerckhove, M. Sudol and J. Gettemans (2010). "TAZ interacts with zonula occludens-1 and -2 proteins in a PDZ-1 dependent manner." FEBS Letters **584**(19): 4175-4180.

Ribeiro, P. S., F. Josue, A. Wepf, M. C. Wehr, O. Rinner, G. Kelly, . . . M. Gstaiger (2010). "Combined functional genomic and proteomic approaches identify a PP2A complex as a negative regulator of Hippo signaling." Molecular Cell **39**(4): 521-534.

Rupp, B. (2009). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. New York, Garland Science.

Salah, Z., G. Melino and R. I. Aqeilan (2011). "Negative regulation of the Hippo pathway by E3 ubiquitin ligase ITCH is sufficient to promote tumorigenicity." Cancer Research **71**(5): 2010-2020.

Salem, M., Y. Mauguen and T. Prange (2010). "Revisiting glutaraldehyde cross-linking: the case of the Arg-Lys intermolecular doublet." Acta Crystallographica Section F: Structural Biology Crystal Communications **66**(Pt 3): 225-228.

Scheel, H. and K. Hofmann (2003). "A novel interaction motif, SARAH, connects three classes of tumor suppressor." Current Biology **13**(23): R899-R900.

Schlegelmilch, K., M. Mohseni, O. Kirak, J. Pruszak, J. R. Rodriguez, D. Zhou, . . . F. D. Camargo (2011). "Yap1 acts downstream of alpha-catenin to control epidermal proliferation." Cell **144**(5): 782-795.

Schneider, A., M. J. Huentelman, J. Kremerskothen, K. Duning, R. Spoelgen and K. Nikolich (2010). "KIBRA: a new gateway to learning and memory?" Frontiers in Aging Neuroscience **2**: 4.

Schwieters, C. D. and G. M. Clore (2001). "The VMD-XPLOR visualization package for NMR structure refinement." Journal of Magnetic Resonance **149**(2): 239-244.

Schwieters, C. D., J. J. Kuszewski and G. M. Clore (2006). "Using Xplor-NIH for NMR molecular structure determination." Progress in Nuclear Magnetic Resonance Spectroscopy **48**(1): 47-62.

Shen, Y., F. Delaglio, G. Cornilescu and A. Bax (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." Journal of Biomolecular NMR **44**(4): 213-223.

Srivastava, A., J. O. MacKay and J. B. Bell (2002). "A Vestigial:Scalloped TEA domain chimera rescues the wing phenotype of a scalloped mutation in Drosophila melanogaster." Genesis **33**(1): 40-47.

Staub, O., S. Dho, P. Henry, J. Correa, T. Ishikawa, J. McGlade and D. Rotin (1996). "WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na+ channel deleted in Liddle's syndrome." EMBO Journal **15**(10): 2371-2380.

Strano, S., E. Munarriz, M. Rossi, L. Castagnoli, Y. Shaul, A. Sacchi, . . . G. Blandino (2001). "Physical interaction with Yes-associated protein enhances p73 transcriptional activity." Journal of Biological Chemistry **276**(18): 15164-15173.

Sudol, M. (1994). "Yes-Associated Protein (Yap65) Is a Proline-Rich Phosphoprotein That Binds to the Sh3 Domain of the Yes Protooncogene Product." Oncogene **9**(8): 2145-2152.

Sudol, M., P. Bork, A. Einbond, K. Kastury, T. Druck, M. Negrini, . . . D. Lehman (1995). "Characterization of the Mammalian Yap (Yes-Associated Protein) Gene and Its Role in Defining a Novel Protein Module, the Ww Domain." Journal of Biological Chemistry **270**(24): 14733-14741.

Sudol, M. and K. F. Harvey (2010). "Modularity in the Hippo signalling pathway." Trends in the Biochemical Sciences.

Sudol, M., K. Sliwa and T. Russo (2001). "Functions of WW domains in the nucleus." FEBS Letters **490**(3): 190-195.

Sutton, R. B., B. A. Davletov, A. M. Berghuis, T. C. Sudhof and S. R. Sprang (1995). "Structure of the first C-2 domain of synaptotagmin 1. A novel Ca^{2+}/phospholipid-binding fold." Cell **80**(6): 929-938.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Research **22**(22): 4673-4680.

Tian, W., J. Yu, D. R. Tomchick, D. Pan and X. Luo (2010). "Structural and functional analysis of the YAP-binding domain of human TEAD2." Proceedings of the National Academy of Sciences of the United States of America **107**(16): 7293-7298.

Tschop, K., A. R. Conery, L. Litovchick, J. A. Decaprio, J. Settleman, E. Harlow and N. Dyson (2011). "A kinase shRNA screen links LATS2 and the pRB tumor suppressor." Genes and Development **25**(8): 814-830.

Udan, R. S., M. Kango-Singh, R. Nolo, C. Y. Tao and G. Halder (2003). "Hippo promotes proliferation arrest and apoptosis in the Salvador/Warts pathway." Nature Cell Biology **5**(10): 914-920.

Valverde, P. (2000). "Cloning, expression, and mapping of hWW45, a novel human WW domain-containing gene." Biochemical and Biophysical Research Communications **276**(3): 990-998.

Varelas, X., B. W. Miller, R. Sopko, S. Y. Song, A. Gregorieff, F. A. Fellouse, . . . L. Attisano (2010). "The Hippo Pathway Regulates Wnt/beta-Catenin Signaling." Developmental Cell **18**(4): 579-591.

Varelas, X., P. Samavarchi-Tehrani, M. Narimatsu, A. Weiss, K. Cockburn, B. G. Larsen, . . . J. L. Wrana (2010). "The Crumbs complex couples cell density sensing to Hippo-dependent control of the TGF-beta-SMAD pathway." Developmental Cell **19**(6): 831-844.

Vassilev, A., K. J. Kaneko, H. J. Shu, Y. M. Zhao and M. L. DePamphilis (2001). "TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm." Genes & Development **15**(10): 1229-1241.

Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel (2000). "Structural basis for phosphoserine-proline recognition by group IV WW domains." Nature Structural Biology **7**(8): 639-643.

Vranken, W. F., W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, P. Llinas, . . . E. D. Laue (2005). "The CCPN data model for NMR spectroscopy: Development of a software pipeline." Proteins-Structure Function and Bioinformatics **59**(4): 687-696.

Webb, C., A. Upadhyay, F. Giuntini, I. Eggleston, M. Furutani-Seiki, R. Ishima and S. Bagby (2011). "Structural features and ligand binding properties of tandem WW domains from YAP and TAZ, nuclear effectors of the Hippo pathway." Biochemistry **50**(16): 3300-3309.

Wiesner, S., G. Stier, M. Sattler and M. J. Macias (2002). "Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40." Journal of Molecular Biology **324**(4): 807-822.

Wishart, D. S., C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, . . . B. D. Sykes (1995). "^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR." Journal of Biomolecular NMR **6**(2): 135-140.

Xiao, L., Y. Chen, M. Ji and J. Dong (2011). "KIBRA regulates Hippo signaling activity via interactions with large tumor suppressor kinases." Journal of Biological Chemistry **286**(10): 7788-7796.

Yagi, R., L. F. Chen, K. Shigesada, Y. Murakami and Y. Ito (1999). "A WW domain-containing Yes-associated protein (YAP) is a novel transcriptional co-activator." EMBO Journal **18**(9): 2551-2562.

Yeates, T. O. (1997). "Detecting and overcoming crystal twinning." Methods in Enzymology **276**: 344-358.

Yoshida, T. (2008). "MCAT elements and the TEF-1 family of transcription factors in muscle development and disease." Arteriosclerosis, Thrombosis, and Vascular Biology **28**(1): 8-17.

Yuan, Z., D. Kim, S. Shu, J. Wu, J. Guo, L. Xiao, . . . J. Q. Cheng (2010). "Phosphoinositide 3-kinase/Akt inhibits MST1-mediated pro-apoptotic signaling through phosphorylation of threonine 120." Journal of Biological Chemistry **285**(6): 3815-3824.

Zaidi, S. K., A. J. Sullivan, R. Medina, Y. Ito, A. J. van Wijnen, J. L. Stein, . . . G. S. Stein (2004). "Tyrosine phosphorylation controls Runx2-mediated subnuclear targeting of YAP to repress transcription." EMBO Journal **23**(4): 790-799.

Zhang, J., J. Y. Ji, M. Yu, M. Overholtzer, G. A. Smolen, R. Wang, . . . D. A. Haber (2009). "YAP-dependent induction of amphiregulin identifies a non-cell-autonomous component of the Hippo pathway." Nat Cell Biol **11**(12): 1444-1450.

Zhang, J., G. A. Smolen and D. A. Haber (2008). "Negative regulation of YAP by LATS1 underscores evolutionary conservation of the Drosophila Hippo pathway." Cancer Research **68**(8): 2789-2794.

Zhao, B., L. Li and K. L. Guan (2010). "Hippo signaling at a glance." Journal of Cell Science **123**(Pt 23): 4001-4006.

Zhao, B., L. Li, Q. Lei and K. L. Guan (2010). "The Hippo-YAP pathway in organ size control and tumorigenesis: an updated version." Genes and Development **24**(9): 862-874.

CHAPTER 8: APPENDICES

Appendix I: Constructs, primers, restriction sites, and vectors

I-a TEAD constructs

Table 7.1 TEAD1 primers

Primer name/ region	Primer	Restriction site	Vectors
TEAD1 outerF/ pre-N-terminal	TAGCTTGGAGAGCCCGACT	Blunt	pBlueScript SK-
TEAD1 innerF/ N-terminal ->	AGAGCCCGACTGCCAAGAT	Blunt	pBlueScript SK-
TEAD1 innerR/ C-terminal <-	tgtTCATTCTTTAACCAGTCTGTAGAT	Blunt	pBlueScript SK-
TEAD1 outerR/ post-C-terminal	ggtgtTCATTCTTTAACCAGTCTGT	Blunt	pBlueScript SK-
Med-TEAD1-for/ (133) pre-YBD	CGGAATTCCGCCAGAAGAAAATCTCGTGAG	EcoRI	Vanderbilt
G246F/ -> YBD	CGGAATTCGGGCCGCTCCATCGGTACTTC	EcoRI	Vanderbilt
Med-TEAD1-rev/ YBD <-	CCGCTCGAGCTATCATTCTTTAACCAGTCTGTAGAT	XhoI	Vanderbilt
TEAD1_13f/ -> DBD	CGGGATCCGTGGACAATGATGCTGAAGG	BamHI	Vanderbilt
TEAD1_92r/ DBDα3 <-	GGAATTCCTATTAAGCTTGGATTTCTCGAACTTTC	EcoRI	Vanderbilt
TEAD1_121r/ DBDα4 <-	GGAATTCCTATTAGTCCATACTTGTAACCTTTAGC	EcoRI	Vanderbilt
TEAD1(1-2)_V12f/ -> DBD	AGGAGATATACCATGGTGGACAATGATGCTGAAGGGGT	N/A - LIC	pOPIN-E
TEAD1_S113r/ DBD <-	GTGATGGTGATGTTTGGAATGAAACTCACGAGATTTTCTTCTTTCAAG	N/A - LIC	pOPIN-E
TEAD1(1-2)_R88f/ -> DBDα4	AGGAGATATACCATGCGAGAAATCCAAGCTGCCATTAAGGTG	N/A - LIC	pOPIN-E
TEAD1(1-2)_T211f/ ->YBD	AGGAGATATACCATGACAGCACCAGCATGGCAGGGC	N/A - LIC	pOPIN-E
TEAD1_E435r/ YBD <-	GTGATGGTGATGTTTTTCTTTAACCAGTCTGTAGATATGATGCTGG	N/A - LIC	pOPIN-E

TEAD1 plasmids:

TEAD1_1-2fl: pBlueScript SK- (full length TEAD1 with DBDα4 cDNA)

TEAD1_1-1fl: pBlueScript SK- (f-l TEAD1 with expected DBD cDNA)

TEAD1 (13-92): pSV281, pBG101 (DBD)

TEAD1 (12-92): pOPIN-E (DBD)

TEAD1 (12-113): pOPIN-E (DBD with α4)

TEAD1 (133-464): pSV281 (YBD and unstructured N-terminal region)

TEAD1 (246-464): pSV281, pBG101 (YBD)

Table 7.2 TEAD2/3 primers

Primer name/ region	Primer	Restriction site	Vectors
Medaka tead3F/ pre-N-terminal	CACCATTGCGTCCAACGA	Blunt	pBlueScript SK-
TEAD3 innerF/ N-terminal ->	GCGTCCAACGAGTGGAGT	Blunt	pBlueScript SK-
TEAD3 innerR/ C-terminal <-	tccctggcaagctgTTAGTC	Blunt	pBlueScript SK-
TEAD3 outerR/ post-C-terminal	AGAAGCTCCCTGGCAAGC	Blunt	pBlueScript SK-
Med-TEAD2-for/ pre-YBD ->	CGGAATTCCGCACGGAAGAAAGTG CGTGAA	EcoRI	Vanderbilt
mTEAD3_V239F/ -> YBD	CGGAATTCGGTTCCAGTATGGCAG GATCG	EcoRI	Vanderbilt
Med-TEAD2-rev/ YBD (464) <-	CCGCTCGAGCTATTAGTCTTTAACA AGTCTGTAGAC	XhoI	Vanderbilt
TEAD3_1f/ -> DBD	CGGGATCCATGGATGGGGACGCAG AGG	BamHI	Vanderbilt
TEAD3_107r_4-1b / DBDα3 <-	GGAATTCCTATTATTGAGCAGACGA CAGCGCT	EcoRI	Vanderbilt
TEAD3_116r_3-1a / DBDα4 <-	GGAATTCCTATTATTTGTCTTTGGA TGCCTGATC	EcoRI	Vanderbilt
TEAD3_M1f/ -> DBD	AAGTTCTGTTTCAGGGCCCGATGGA TGGGGACGCAGAGGGG	N/A - LIC	pOPIN-F
TEAD3_G82r/ DBDα3 <-	ATGGTCTAGAAAGCTTTAGCCCGCC TGGTATTCACGCAC	N/A - LIC	pOPIN-F
TEAD3_K105r/ DBD <-	ATGGTCTAGAAAGCTTTACTTTAGC TTTGACTGGATCTCGCGAGA	N/A - LIC	pOPIN-F
TEAD3_E78f/ -> DBDα4	AAGTTCTGTTTCAGGGCCCGGAATA CCAGGCCGGCATAAAG	N/A - LIC	pOPIN-F
TEAD3_T201f/ ->YBD	AAGTTCTGTTTCAGGGCCCGACCGC AGTTCCAGTATGGCAGGAT	N/A - LIC	pOPIN-F
TEAD3_D247r/ YBD <-	ATGGTCTAGAAAGCTTTAGTCTTTA ACAAGTCTGTAGACGTGCTCTTG	N/A - LIC	pOPIN-F

TEAD2/3 plasmids:

TEAD3_3-1a_fl: pBlueScript SK- (full length TEAD2/3 with DBDα4 cDNA)

TEAD3_4-1b_fl: pBlueScript SK- (f-l TEAD2/3 with expected DBD cDNA)

TEAD3 (1-82): pOPIN-F/-J (DBD)

TEAD3 (1-105): pOPIN-F/-J (DBD with α4)

TEAD3 (1-107): pSV281 (DBD with α4)

TEAD2 (239-462): pSV281, pBG101 (YBD)

TEAD3 (201-427): pOPIN-F/-J (YBD)

Table 7.3 TEAD4 primers

Primer name/ region	Primer	Restriction site	Vectors
Medaka tead4F/ pre-N-terminal	AGAGGTTCCCTCTCCATCGT	Blunt	pBlueScript SK-
TEAD4 innerF/ N-terminal ->	CATCGTCCCACCACGTCT	Blunt	pBlueScript SK-
TEAD4 innerR/ C-terminal <-	TCAGTCTTTGACCAGCCTGT	Blunt	pBlueScript SK-
Medaka TEAD4R/ post-C-terminal	CACCAGTGCCTCTGTGTCAG	Blunt	pBlueScript SK-
TEAD4_Y222f/ -> YBD	CGGAATTCGTATGCTATGCAAGCG ACTGG	EcoRI	Vanderbilt
TEAD4_D470r/ YBD <-	CCGCTCGAGCTATCAGTCTTTGACC AGCCTGTAAAT	XhoI	Vanderbilt
TEAD4_194f/ -> YBD	CGGGATCCGGTAGAAGTATTGCCA GCTC	BamHI	Vanderbilt
TEAD4_216f/ -> YBD	CGGGATCCGACCCAGAGACTTTCA ACAAG	BamHI	Vanderbilt
TEAD4_239f/ -> YBD	CGGGATCCTATCTGGAGTCGGTGG ACAT	BamHI	Vanderbilt
TEAD4_A2for/ -> DBD	AAGTTCTGTTTCAGGGCCCGGCAGA GGGTGTGTGGAGCCCT	N/A - LIC	pOPIN-F/J
TEAD4_Q77rev/ DBD <-	ATGGTCTAGAAAGCTTTACTGGATC TCCCGGGCCTTACG	N/A - LIC	pOPIN-F/J
TEAD4_M171for/ ->YBD	AAGTTCTGTTTCAGGGCCCGATGTC TCCTGGTGCCCCCCCT	N/A - LIC	pOPIN-F/J
TEAD4_D398rev/ YBD <-	ATGGTCTAGAAAGCTTTAGTCTTTG ACCAGCCTGTAAATATGGTGC	N/A - LIC	pOPIN-F/J

TEAD4 plasmids:

TEAD4-1 fl: pBlueScript SK- (TEAD4 full length cDNA)

TEAD4 (222-470): pBG100/101, pSV281 (YBD)

TEAD4 (194-470): pBG100/101, pSV281 (YBD and unstructured N-terminus)

TEAD4 (171-398): pOPIN-F/-J (YBD)

I-b YAP constructs

Table 7.4 YAP primers

Primer name/ region	Primer	Restriction site	Vectors
YAP_M1f/ N-terminal ->	CGGAATTCCATGGATCCGAGCCAGCACAAC	EcoRI	Vanderbilt
YAP_V10f/ TBD ->	CGGAATTCCGTCGGACACCAGATTGTTCAC	EcoRI	Vanderbilt
YAP_M105f/ pre-WW1 ->	CGGAATTCCATGGCCTCAGCAGGCGCCTC	EcoRI	Vanderbilt
YAP_S119f/ WW1 ->	CGGAATTCCTCCTACGAGATTCCCGAC	EcoRI	Vanderbilt
YAP_D124f/ short WW1 ->	CGGAATTCCGACGACGTGCCCCTG	EcoRI	Vanderbilt
YAP S171f/ pre-WW2 ->	CGGAATTCCAGTACCGTTCCAGTTCAGCAG	EcoRI	Vanderbilt
YAP S182f/ WW2 ->	CGGAATTCCAGCCCAGCTAGTGGAC	EcoRI	Vanderbilt
YAP S65r/ short TBD <-	CCCAAGCTTCTATTAGGAATGGGACTTGGGCTC	HindIII	Vanderbilt
YAP L94r/ TBD <-	CCCAAGCTTCTATTACAGCTGCAGGGAAGCTGG	HindIII	Vanderbilt
YAP G109r/ post-TBD <-	CCCAAGCTTCTATTAGCCTGCTGAGGCCATCC	HindIII	Vanderbilt
YAP L162r/ WW1 <-	CCCAAGCTTCTATTAGAGCAGGGCTTTGCG	HindIII	Vanderbilt
YAP T168r/ post-WW1 <-	CCCAAGCTTCTATTAGGTAGCCTGGTTCAACTG	HindIII	Vanderbilt
YAP R222r/ WW2 <-	CCCAAGCTTCTATTAGCGAGTCTCAAGTCG	HindIII	Vanderbilt
YAP A240r/ post-WW2 <-	CCCAAGCTTCTATTAAGCGCTTTGGGTGATCCT	HindIII	Vanderbilt
YAP L440r/ C-terminal <-	CCCAAGCTTCTATTATAACCATGTGAGGAAGCTCTCCTT	HindIII	Vanderbilt
YAP_V10F pOPINF/TBD->	AAGTTCTGTTTCAGGGCCCGGTCGGACACCAGATTGTTCACGTT	N/A-LIC	pOPIN-F/J
YAP_Q67R pOPINF/TBD<-	ATGGTCTAGAAAGCTTTATTGTCTGGAATGGGACTTGGGCTC	N/A-LIC	pOPIN-F/J
YAP_G100R pOPINF/TBD<-	ATGGTCTAGAAAGCTTTAGCCACCCGAAACGGCGCC	N/A-LIC	pOPIN-F/J

YAP plasmids:

YAP (10-65): pSV281, pBG101	(minimal TBD)
YAP (10-94): pSV281, pBG101	(TBD)
YAP (1-109): pSV281/282, pBG101	(TBD)
YAP (1-100): pOPIN-F/-J	(TBD)
YAP (10-67): pOPIN-F/-J	(minimal TBD)
YAP (1-168): pSV281, pBG100	(TBD-WW1)
YAP (1-234): pSV281, pBG100	(TBD-WW1-WW2)
YAP (105-168): pSV281, pBG101	(long WW1 with unstructured termini)
YAP (119-162): pSV281, pBG101	(WW1)
YAP (124-162): pSV281, pBG101	(short, partially unfolded WW1)
YAP (105-234): pSV281, pBG101	(long WW1-long WW2)
YAP (119-222): pSV281, pBG101	(WW1-WW2)
YAP (124-222): pSV281, pBG101	(short WW1-WW2)
YAP (171-234): pSV281, pBG100/101	(long WW2)
YAP (182-222): pSV281, pBG101	(WW2)

I-c TAZ constructs

Table 7.5 TAZ primers

Primer name/ region	Primer	Restriction site	Vectors
TAZ 3F/ pre-N-terminal ->	TGGAGATTTGCTTCGACTCTG	Blunt	pBluescript SK-
TAZ 4F/ N-terminal ->	GACTCTGACTTCTGGCGACAT	Blunt	pBluescript SK-
TAZ1 innerR/ C-terminal <-	GGTGAGCAGATTGTCCCTGT	Blunt	pBluescript SK-
TAZ 3R/ post-C-terminal <-	GCCTATAGCCAGGTGAGCAG	Blunt	pBluescript SK-
TAZ G7f/ TBD ->	CGGAATTCCGGACGGCAGGTCGTGCAC	EcoRI	Vanderbilt
TAZ A103f/ WW1 ->	CGGAATTCCGCGACATCGATCATACCTGA	EcoRI	Vanderbilt
TAZ N177f/ WW2 ->	CGGAATTCCAACCCGGAGTCAGGTCCAC	EcoRI	Vanderbilt
TAZ L147r/ WW1 <-	CCCAAGCTTCTATTAGAGCTGCGAGAGACGAGG	HindIII	Vanderbilt
TAZ M218r/ WW2 <-	CCCAAGCTTCTATTACATTTTCTGAGCTAGACGCG	HindIII	Vanderbilt
TAZ_R5f_pOPIN/ TBD ->	AAGTTCTGTTTCAGGGCCCGCGGAGCGGACGGCAGGTC	N/A-LIC	pOPIN-F/J
TAZ_R63r_pOPIN/ TBD <-	ATGGTCTAGAAAGCTTTATCTAGAGTGTCCGCGGGGCTC	N/A-LIC	pOPIN-F/J
TAZ_T96r_pOPIN/ TBD <-	ATGGTCTAGAAAGCTTTAGGTGGAGAGGGAGTTGACGGG	N/A-LIC	pOPIN-F/J

TAZ plasmids:

TAZ (5-63): pOPIN-F/-J	(minimal TBD)
TAZ (5-96): pOPIN-F/-J	(TBD)
TAZ (7-147): pSV281, pBG100/101	(TBD-WW1)
TAZ (7-218): pSV281, pBG100/101	(TBD-WW1-WW2)
TAZ (103-147): pSV281, pBG100/101	(WW1)
TAZ (103-218): pSV281, pBG100/101	(WW1-WW2)
TAZ (177-218): pSV281, pBG100/101	(WW2)

I-d Sav constructs

Table 7.6 Sav primers

Primer name/ region	Primer	Restriction site	Vectors
DrosSav_E421f/ WW1 ->	CGGGATCCGAGGAGCTGCCTCTGCCG	BamHI	Vanderbilt
DrosSav_L455f/ WW2 ->	CGGGATCCTTGGAGCGCGAAGGTCTG	BamHI	Vanderbilt
DrosSav_R457r/ WW1 <-	GGAATTCTTATTAGCGCTCCAACGGAT GATTC	EcoRI	Vanderbilt
DrosSav_P489r/ WW2 <-	GGAATTCCTATTATGGATGCTGACGTTG GCTC	EcoRI	Vanderbilt
DrosSav_Q608r/ SARAH <-	GGAATTCCTATCACTGGTTTTGGTTCTG GTTC	EcoRI	Vanderbilt
hWW45_E197f/ WW1 ->	CGGGATCCGAAGATTTACCCCTTCCTCC	BamHI	Vanderbilt
hWW45_P265r/ WW2 <-	GGAATTCCTATTAGGGATGCCTGTATTG GGC	EcoRI	Vanderbilt
hWW45_G380r/ SARAH <-	GGAATTCCTATCATCCATGTTGTTGGGC ATACC	EcoRI	Vanderbilt

***Drosophila* Sav plasmids:**

dSav WW1: pSV281, pBG101 (E421-R457)

dSav WW1-WW2: pSV281, pBG101 (E421-P489)

dSav WW1–WW2–SARAH: pSV281, pBG101 (E421-Q608)

dSav WW2: pSV281, pBG101 (L455-P489)

dSav WW2-SARAH: pSV281, pBG101 (L455-Q608)

Human Sav1 (WW45) plasmids:

hSav WW1-WW2: pSV281, pBG101 (E197-P265)

hSav WW1-WW2-SARAH: pSV281, pBG101 (E197-G380)

I-e Kibra constructs

Table 7.7 Kibra primers

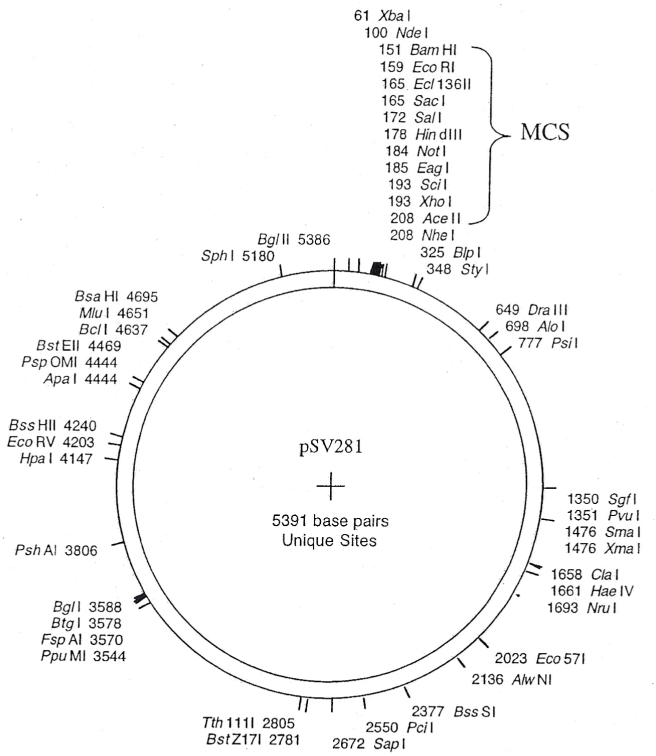
Primer name/ region	Primer	Restriction site	Vectors
Kibra_E5f/ WW1 ->	CGGAATTCCGAGCTGCCCCTGCCGGA	EcoRI	Vanderbilt
Kibra_A48f/ WW2 ->	CGGAATTCCGCTGACTGCATTAGTGATG A	EcoRI	Vanderbilt
Kibra_K43r/ WW1 <-	CCCAAGCTTCTATTATTTGGTGTACCTGT CCCG	HindIII	Vanderbilt
Kibra_Q87r/ WW2 <-	CCCAAGCTTCTATTATTGTACTCGAGGAT CCTCAAT	HindIII	Vanderbilt
Kibra_K97r/ WW2(l) <-	CCCAAGCTTCTATTACTTCAGCATATGTT CCTGCT	HindIII	Vanderbilt

Kibra plasmids:

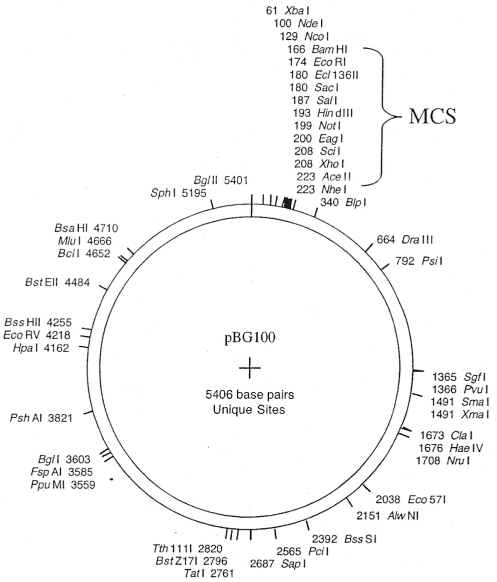
Kibra (5-87): pSV281, pBG101	(WW1-WW2)
Kibra (5-97): pSV281, pBG101	(WW1-WW2 long)
Kibra WW1+WW2 WT: pGEX42	(from J. Kremerskothen)
Kibra WW1(P37A)+WW2(P84A): pGEX42	(from J. Kremerskothen)
Kibra WW1(P37A)+WW2: pGEX42	(from J. Kremerskothen)
Kibra WW1 WT: pGEX-2TK-KG	(from J. Kremerskothen)
Kibra WW1(P37A): pGEX-2TK-KG	(from J. Kremerskothen)
Kibra WW2 WT: pGEX-2TK-KG	(from J. Kremerskothen)
Kibra WW2(P84A): pGEX-2TK-KG	(from J. Kremerskothen)

I-f Vanderbilt vectors

The Vanderbilt vector maps for pSV281, pBG100, and pBG101 used during this project are shown below.

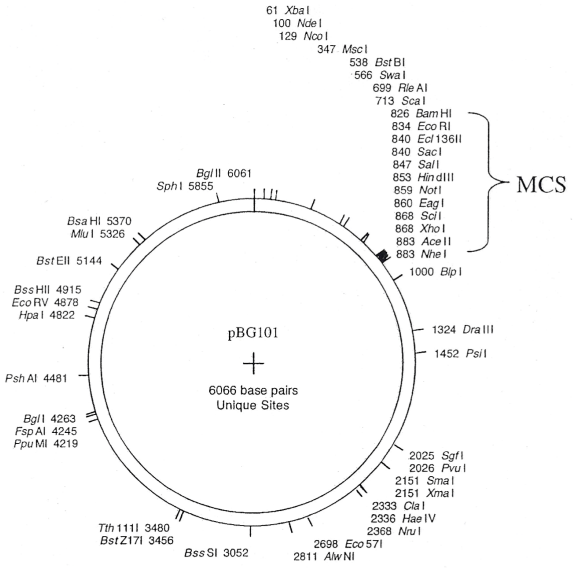



```
                                                                  XbaI
  1 cgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgttt
                          NdeI                                                BamHI
 80 aactttaagaaggagatataCATATGggttcttctcaccatcaccatcaccatgaaaacctgtacttccagGGATCCcc
                         1▶ M  G  S  S  H  H  H  H  H  H  E  N  L  Y  F  Q▲G  S  P
                                  EagI                  His Tag                 TEV Protease Site
     EcoRI  SacI     SalI    HindIII NotI        XhoI          NheI
159 GAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGatcaaacggGCTAGCcagccagaactcgccccggaagac
 19▶  N  S  S  S  V  D  K  L  A  A  A  L  E  I  K  R  A  S  Q  P  E  L  A  P  E  D
238 cccgaggatgtcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagct...
 46▶ P  E  D  V  E  H  H  H  H  H  H
```




```
                                                                  XbaI
1   cgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaacttta
                  NdeI                          NcoI                                  BamHI
87  agaaggagatataCATATGggttcttctcaccatcaccatcaccatggcagcagcctggaagttctgttccaggggcccGGATCCc
                 1▶ M  G  S  S  H  H  H  H  H  H  G  S  S  L  E  V  L  F  Q▲G  P  G  S
                                  His Tag                          3C Protease site
                                      EagI
     EcoRI  SacI     SalI    HindIII NotI       XhoI            NheI
173  cGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGatcaaacgggctagccagccagaactcgccccggaagaccccgag
 24▶P  N  S  S  S  V  D  K  L  A  A  A  L  E  I  K  R  A  S  Q  P  E  L  A  P  E  D  P  E

259  gatgtcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgc...
 53▶ D  V  E  H  H  H  H  H  H
```




```
                                                                  XbaI
1   cgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaaga
                  NdeI                          NcoI                          EcoNI
90  aggagatataCATATGggttcttctcaccatcaccatcaccatggttcttctatgtccctatactaggttattgg  ... 639 bp ...
                 1▶ M  G  S  S  H  H  H  H  H  H  G  S  S  M  S  P  I  L  G  Y  W  ... 213 aa ...
                                  His Tag                          GST Tag
                                                                      EagI
                          BamHI    EcoRI  SacI    SalI    HindIII NotI       XhoI            NheI
802  ctggaagttctgttccaggggcccGGATCCccGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGatcaaacggGCTAGCca
234▶ L  E  V  L  F  Q▲G  P  G  S  P  N  S  S  S  V  D  K  L  A  A  A  L  E  I  K  R  A  S  Q
     3C Protease site

891  gccagaactcgccccggaagaccccgaggatgtcgagcaccaccaccaccaccactgagatccggctgctaa...
263▶ P  E  L  A  P  E  D  P  E  D  V  E  H  H  H  H  H  H
```

I-g pOPIN vectors

Details of pOPIN vectors used for ligation independent cloning are shown below. The vectors used in Chapter 3 have been highlighted with a red asterisk.

pOPIN Vector Characteristics, Derived Affinity Tags and Primer Extensions

Vector	Fusion Tag	Parent Vector/ Antibiotic resistance	Promoters/ Baculoviral Recombination Sites	Forward Primer Extension	Reverse Primer Extension
pOPINA	...KHHHHHH tag	pET28a/Kanamycin	T7lacO	AGGAGATATACCATG	GTGGTGGTGGTGTTT
pOPINB	MGSSHHHHHHSSGLEVLFQ**U**GP ...tag	pET28a/Kanamycin	T7lacO	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINE	...KHHHHHH tag	pTriEx2/Ampicillin	T7lacO, CMV enhancer and ß-actin promoter, p10 promoter /lef-2 and 1629 baculo elements.	AGGAGATATACCATG	GTGATGGTGATGTTT†
pOPINF	MAHHHHHHSSGLEVLFQ**U**GP... tag	pTriEx2/Ampicillin	T7lacO, CMV enhancer and ß-actin promoter, p10 promoter /lef-2 and 1629 baculo elements.	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPING	MGILPSPGMPALLSLVSLLSVLL MGCVA**U**ETG... cleavable secretion leader and ...KHHHHHH tags	pTriEx2/Ampicillin	(T7lacO-not used), CMV enhancer and ß-actin promoter, p10 promoter /lef-2 and 1629 baculo elements.	GCGTAGCTGAAACCGGC	GTGATGGTGATGTTT
pOPINH	MGILPSPGMPALLSLVSLLSVLLMG CVA**U**ETMAHHHHHHSSGLEVLFQ **U**GP..... cleavable secretion leader and cleavable N-his tag	pTriEx2/Ampicillin	(T7lacO-not used), CMV enhancer and ß-actin promoter, p10 promoter /lef-2 and 1629 baculo elements.	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINI	MAHHHHHHSSG... tag	pTriEx2/Ampicillin	T7lacO, CMV enhancer and ß-actin promoter, p10 promoter /lef-2 and 1629 baculo elements..	ACCATCACAGCAGCGGC	ATGGTCTAGAAAGCTTTA
pOPINJ	MAHHHHHH-GST-SSGLEVLFQ**U**GP... tag	pTriEx2/Ampicillin	T7lacO, CMV enhancer and ß-actin promoter, p10 promoter /lef-2 and 1629 baculo elements..	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINK	MAHHHHHH-GST-SSGLEVLFQ**U**GP... tag	pET28a/Kanamycin	T7lacO	AAGTTCTGTTTCAGGGCCCG AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA ATGGTCTAGAAAGCTTTA
pOPINM	MAHHHHHHSSG-MBP-LEVLFQ**U**GP... tag	pTriEx2/Ampicillin	T7lacO, CMV enhancer and ß-actin promoter, p10 promoter /lef-2 and 1629 baculo elements.	AAGTTCTGTTTCAGGGCCCG	ATGGTCTAGAAAGCTTTA
pOPINS	MGSSHHHHHH-SUMO**U** ... tag	pET28a/Kanamycin	T7lacO	GCGAACAGATCGGTGGT	ATGGTCTAGAAAGCTTTA
pOPINTTG	MGILPSPGMPALLSLVSLLSVLL MGCVA**U**ETG... cleavable secretion leader and ...KHHHHHH tags	pTT3	CMV/Adenovirus ML enhancer/tripartite leader and EBNA ori for HEK-EBNA cells	GCGTAGCTGAAACCGGC	GTGATGGTGATGTTT
pOPINTTI	MAHHHHHHSSG... tag	pTT3	CMV/Adenovirus ML enhancer/tripartite leader and EBNA ori for HEK-EBNA cells.	ATCATCACAGCAGCGGC	ATGGTCTAGAAAGCTTTA

Key to table:
ATG-These primers contain an ATG for domain expression-the ATG may be omitted from the gene-specific forward primer sequence.
TTA-These primers contain a TAA for domain expression-the stop codon may be omitted from the gene-specific reverse primer sequence.
U-denotes specific protease cleavage site for 3C protease.
U-denotes specific protease cleavage site for signal peptidase.
U-denotes specific protease cleavage site for SUMO protease.
pOPIN- a PCR product containing these primer pairs may be cloned into any of these vectors.
Italics-denotes a potentially solubilising fusion partner.

Appendix II: DNA sequencing results and protein sequences

Only the longest sequenced constructs for each protein are included here. All shorter constructs were sequenced successfully and encompass shorter sections of the sequences shown below. Grey sequences represent a His-tag. Rare codon analysis was carried out using the UCLA Rare codon calculator (RaCC) for TEAD proteins (Tables 7.8-7.10) due to difficulties in protein expression and purification.

II-a Medaka TEAD1 full-length sequences

TEAD1_1-2fl (TEAD1_1-1fl lacks the DBDα4 and the penultimate K residue):

```
atggagaggatgagtgattcagcagacaagcctgtggacaatgatgctgaaggggtgtgg
 M  E  R  M  S  D  S  A  D  K  P  V  D  N  D  A  E  G  V  W
agccccgacattgaacaaagctttcaagaagctctggccatctaccctccatgtggtcga
 S  P  D  I  E  Q  S  F  Q  E  A  L  A  I  Y  P  P  C  G  R
aggaagatcatcctttctgatgagggaaagatgtatggtcgaaatgaactgatagccaga
 R  K  I  I  L  S  D  E  G  K  M  Y  G  R  N  E  L  I  A  R
tacatcaagctccgtaccggaaagacgaggacccggaagcaggtgtctagtcacatacag
 Y  I  K  L  R  T  G  K  T  R  T  R  K  Q  V  S  S  H  I  Q
gtcttagcaagaaagaaagttcgagaaatccaagctgccattaaggtgtctagtcacatt
 V  L  A  R  K  K  V  R  E  I  Q  A  A  I  K  V  S  S  H  I
caggttcttgccagaagaaaatctcgtgagtttcattccaagctaaaggttacaagtatg
 Q  V  L  A  R  R  K  S  R  E  F  H  S  K  L  K  V  T  S  M
gaccaagctgtgaaagacaaagccctccagagcatggcctccatgtcctcggctcagatc
 D  Q  A  V  K  D  K  A  L  Q  S  M  A  S  M  S  S  A  Q  I
gtctctgccacggcgattcacaataagctgggtctgccgggcatccctcgtccagccttt
 V  S  A  T  A  I  H  N  K  L  G  L  P  G  I  P  R  P  A  F
cctggagcagggttatggcagggtatgatatcagctggtcaaccaggatcctcacaagac
 P  G  A  G  L  W  Q  G  M  I  S  A  G  Q  P  G  S  S  Q  D
attaagcctttcacccagcaaccctatcccatccagccagcagtcacaacaaccatttca
 I  K  P  F  T  Q  Q  P  Y  P  I  Q  P  A  V  T  T  T  I  S
agttatgatccctcagcggctccgacacccacagcaccagcatggcagggccgctccatc
 S  Y  D  P  S  A  A  P  T  P  T  A  P  A  W  Q  G  R  S  I
ggtacttctaaactccggctggtggagttctccgccttcttggaacaacagagagaccca
 G  T  S  K  L  R  L  V  E  F  S  A  F  L  E  Q  Q  R  D  P
gactcttaccataagcatctgtttgtacatattggacagataaatcattcctacagtgac
 D  S  Y  H  K  H  L  F  V  H  I  G  Q  I  N  H  S  Y  S  D
gccctgctggagtcagtggacattcgtcagatttatgacaaattccctgaaaagaaaggg
 A  L  L  E  S  V  D  I  R  Q  I  Y  D  K  F  P  E  K  K  G
ggcctgaaagagctttatggaagaggaccccagaactccttcttcctcatcaaattctgg
 G  L  K  E  L  Y  G  R  G  P  Q  N  S  F  F  L  I  K  F  W
gctgacttgaactgcaatattcaggatgagaccggatctttctacagagtcaccagtcag
 A  D  L  N  C  N  I  Q  D  E  T  G  S  F  Y  R  V  T  S  Q
tatgaaagttcagagaacatgaccattacatgctcaacaaaggtctgctcctttggcaaa
 Y  E  S  S  E  N  M  T  I  T  C  S  T  K  V  C  S  F  G  K
caagtggtggagaaggttgagacggaatatgctcgctttgagaacggacgattcgtctac
 Q  V  V  E  K  V  E  T  E  Y  A  R  F  E  N  G  R  F  V  Y
cggataagtcggtctccgatgtgtgaatacatgatcaacttcatccacaagcttaagcat
 R  I  S  R  S  P  M  C  E  Y  M  I  N  F  I  H  K  L  K  H
ctgcccgagaaatacatgatgaacagcgtgctggagaacttcaccatcttattggtggta
 L  P  E  K  Y  M  M  N  S  V  L  E  N  F  T  I  L  L  V  V
acaaacagggacacccaggagacgctgctgtgcatggcgtgtgtgtttgaagtctcaaac
 T  N  R  D  T  Q  E  T  L  L  C  M  A  C  V  F  E  V  S  N
agtgagcacggggcccagcatcatatctacagactggttaaagaatga
 S  E  H  G  A  Q  H  H  I  Y  R  L  V  K  E  -
```

Table 7.8 Rare codon analysis of TEAD1 (RaCC)	
Rare codons	Rare codon positions (rare codon doublets are **highlighted**)
Arg: agg/aga/cga	3, **40, 41**, 54, 60, 70, 84, 88, **105, 106**, 238, 288, 316, 357, 403, 431
Leu: cta	115
Ile: ata	58, 79, 170, 254, 362
Pro: ccc	22, 188, 190, 204, 210, 290, 382

II-b Medaka TEAD2/3 full-length sequences

TEAD 3-1a fl:

```
atggatggggacgcagaggggtttggagcccagatatcgagcagagcttccaggaggcc
 M  D  G  D  A  E  G  V  W  S  P  D  I  E  Q  S  F  Q  E  A
cttgccatctaccctccatgtggtaggaggaaaatcatactgtccgacgagggaaagatg
 L  A  I  Y  P  P  C  G  R  R  K  I  I  L  S  D  E  G  K  M
tacggtcgtaatgaactgattgcaagatatatcaagctgaggacaggcaaaacccgcaca
 Y  G  R  N  E  L  I  A  R  Y  I  K  L  R  T  G  K  T  R  T
agaaagcaggtgtctagtcacatacaggtgttagcacggaagaaagtgcgtgaataccag
 R  K  Q  V  S  S  H  I  Q  V  L  A  R  K  K  V  R  E  Y  Q
gcgggcataaaggtttctagccacttgcaggttctcgcccggagaaaatctcgcgagatc
 A  G  I  K  V  S  S  H  L  Q  V  L  A  R  R  K  S  R  E  I
cagtcaaagctaaaggcaatgaatttggatcaggcatccaaagacaaagctctacagaac
 Q  S  K  L  K  A  M  N  L  D  Q  A  S  K  D  K  A  L  Q  N
atggcagcgctgtcgtctgctcaaatcgtctccccgagtatgataaagaatcatcttcca
 M  A  A  L  S  S  A  Q  I  V  S  P  S  M  I  K  N  H  L  P
ccaatgtctccagccccatatcagcctcccagattctggcctggtccaatctctgtgcag
 P  M  S  P  A  P  Y  Q  P  P  R  F  W  P  G  P  I  S  V  Q
cctggaccctctcaggacataaaaccatttgtacagcccccataccaaagccttcaaggc
 P  G  P  S  Q  D  I  K  P  F  V  Q  P  P  Y  Q  S  L  Q  G
cccgtacaaccagctataccagcctatgagcctttagcgcctcctccggcaccaacagcc
 P  V  Q  P  A  I  P  A  Y  E  P  L  A  P  P  P  A  P  T  A
accgcagttccagtatggcaggatcgaacgattgcttcctccaagctgcggatgttggag
 T  A  V  P  V  W  Q  D  R  T  I  A  S  S  K  L  R  M  L  E
tactctgcctttatggaggtccagagagaccccgacaactacagcaaacacttgtttgtc
 Y  S  A  F  M  E  V  Q  R  D  P  D  N  Y  S  K  H  L  F  V
cacattggacagacgaatccctcatacagtgatccactccttgaagctgtggacatccgg
 H  I  G  Q  T  N  P  S  Y  S  D  P  L  L  E  A  V  D  I  R
cagatttacgacaagttccctgagaagaaaggaggcctaaaagagctttatgagaaaggc
 Q  I  Y  D  K  F  P  E  K  K  G  G  L  K  E  L  Y  E  K  G
ccccagaatgctttcttcttagtaaagttctgggctgacctgaacagcagcgggatgcca
 P  Q  N  A  F  F  L  V  K  F  W  A  D  L  N  S  S  G  M  P
gatgggcccggctcattctacggggtcagcagccagtacagcagcgccgaaaatatgaca
 D  G  P  G  S  F  Y  G  V  S  S  Q  Y  S  S  A  E  N  M  T
atcactgtttccaccaaggtgtgctcctttggcaagcaggttgtcgagaaagttgagaca
 I  T  V  S  T  K  V  C  S  F  G  K  Q  V  V  E  K  V  E  T
gagtacgcccggctggagggaggaaagtgcgtttacaggatccaccgctcgccaatgtgc
 E  Y  A  R  L  E  G  G  K  C  V  Y  R  I  H  R  S  P  M  C
gagtacatgatcaactttatccacaagctcaaacacctgccggaaaaatacatgatgaac
 E  Y  M  I  N  F  I  H  K  L  K  H  L  P  E  K  Y  M  M  N
agtgttctggaaaactttacaattctacaggtggtgacaaatcgcgacacccaggagacc
 S  V  L  E  N  F  T  I  L  Q  V  V  T  N  R  D  T  Q  E  T
ttgctttgtatagcatttgtgtttgaggtttccaatagtgagcacggagctcaatatcac
 L  L  C  I  A  F  V  F  E  V  S  N  S  E  H  G  A  Q  Y  H
gtctacagacttgttaaagactaa
 V  Y  R  L  V  K  D  -
```

Table 7.9 Rare codon analysis of TEAD2/3 (RaCC)	
Rare codons	Rare codon positions (rare codon doublets are **highlighted**)
Arg: agg/aga/cga	**29, 30**, 49, 54, 61, 95, 151, 209, 229, 353, 423
Leu: cta	104, 118, 273, 389
Ile: ata	33, 68, 83, 135, 167, 186, 404
Pro: ccc	150, 163, 173, 181, 231, 247, 281, 303

II-c Medaka TEAD4 full-length sequences

TEAD4-1fl (*a start codon could not be identified for medaka TEAD4)*

```
aagcttgatcatcgtcccaccacgtctgcggagtgggtgccggctccctcccccaggggt
 K  L  D  H  R  P  T  T  S  A  E  W  V  P  A  P  S  P  R  G
gggagtgaggagctgggggatggacttgacaaaccccctggagaacgatgcagagggtgtg
 G  S  E  E  L  G  D  G  L  D  K  P  L  E  N  D  A  E  G  V
tggagccctgacattgagcagagttttcaggaggctctggccatctatccaccctgcggc
 W  S  P  D  I  E  Q  S  F  Q  E  A  L  A  I  Y  P  P  C  G
cgccgcaagatcatcctctctgatgagggcaagatgtacggacggaatgagctgattgca
 R  R  K  I  I  L  S  D  E  G  K  M  Y  G  R  N  E  L  I  A
cggtacatcaaacttcgtacggggaagacacgaaccaggaagcaggtatctagtcatatc
 R  Y  I  K  L  R  T  G  K  T  R  T  R  K  Q  V  S  S  H  I
caggtccttgcccggcgtaaggcccgggagatccaggtgaagctgaaggaccaagctgcc
 Q  V  L  A  R  R  K  A  R  E  I  Q  V  K  L  K  D  Q  A  A
aaggacaaggctctccagagtatggccaccatgtcctcagcccagatcatctcccccaca
 K  D  K  A  L  Q  S  M  A  T  M  S  S  A  Q  I  I  S  P  T
gccttccagaacaagatgactctccagggcctctctagaccggtttaccccactgctggt
 A  F  Q  N  K  M  T  L  Q  G  L  S  R  P  V  Y  P  T  A  G
gggttctggcatggggctcttcccggacagccaggaagccatgaagacattaagcccttc
 G  F  W  H  G  A  L  P  G  Q  P  G  S  H  E  D  I  K  P  F
tcccaccagagttatgctatgcaagcgactggcccggccccaatagcaggctatgaaagt
 S  H  Q  S  Y  A  M  Q  A  T  G  P  A  P  I  A  G  Y  E  S
acagcagggctgtccatgtctcctggtgcccccccttggcaaggtagaagtattgccagc
 T  A  G  L  S  M  S  P  G  A  P  P  W  Q  G  R  S  I  A  S
tccaagctgcgaatgctggagttctccgccttcttggagcaacctcaagacccagagact
 S  K  L  R  M  L  E  F  S  A  F  L  E  Q  P  Q  D  P  E  T
ttcaacaagcacttatttgttcacatcggacagtccaacccgacctacagcgacccctat
 F  N  K  H  L  F  V  H  I  G  Q  S  N  P  T  Y  S  D  P  Y
ctggagtcggtggacatcaggcagatctatgacaaattccctgagaagaaaggggcctc
 L  E  S  V  D  I  R  Q  I  Y  D  K  F  P  E  K  K  G  G  L
aaggagctttcgacaaagggccccacaatgctttctttcttgtcaagttttgggcggac
 K  E  L  F  D  K  G  P  H  N  A  F  F  L  V  K  F  W  A  D
ctgagtgtaaacctgcaagacgacagcaacttcttttatggcgtgtccagccagtatgag
 L  S  V  N  L  Q  D  D  S  N  F  F  Y  G  V  S  S  Q  Y  E
agctctgacaacatggttatcacctcatccaccaaagtgtgctcctttggcaaacaggtt
 S  S  D  N  M  V  I  T  S  S  T  K  V  C  S  F  G  K  Q  V
gttgagaaagtggagacggagtacgcgcgcttcgagaacgggcgctacgtgttccggatc
 V  E  K  V  E  T  E  Y  A  R  F  E  N  G  R  Y  V  F  R  I
cacagatccccgttgtgtgaatacatgatcaacttcattcacaagctgaagcatctgccg
 H  R  S  P  L  C  E  Y  M  I  N  F  I  H  K  L  K  H  L  P
gagaagtacatgatgaacagcgtactggaaaacttcaccatcctacaggtggtgactaac
 E  K  Y  M  M  N  S  V  L  E  N  F  T  I  L  Q  V  V  T  N
agggacacactggagaccctgctgtgcatagcatacgtcttcgaggtgtccaccagcgag
 R  D  T  L  E  T  L  L  C  I  A  Y  V  F  E  V  S  T  S  E
cacggcgcacagcaccatatttacaggctggtcaaagactga
 H  G  A  Q  H  H  I  Y  R  L  V  K  D  -
```

Table 7.10 Rare codon analysis of TEAD4 (RaCC)	
Rare codons	Rare codon positions
Arg: agg/aga/cga	13, 85, 87, 147, 210, 218, 261, 356, 395, 423
Leu: cta	389
Ile: ata	189, 404
Pro: ccc	10, 12, 26, 52, 133, 151, 162, 173, 205, 253, 282

II-d Medaka YAP 1-234 (TBD-WW1-WW2) sequences

```
atggatccgagccagcacaaccctcctgtcggacaccagatcgttcacgttcggggcgac
 M  D  P  S  Q  H  N  P  P  V  G  H  Q  I  V  H  V  R  G  D
tcggaaacggacctggaggcgcttttcaacgccgtgatgaacccaaagggcgccgtcgtg
 S  E  T  D  L  E  A  L  F  N  A  V  M  N  P  K  G  A  V  V
ccccagtccgtgcccatgaggatgaggaagctgccggactccttcttcaagcccccggag
 P  Q  S  V  P  M  R  M  R  K  L  P  D  S  F  F  K  P  P  E
cccaagtcccattccagacaagctagcactgatgctgggagtggcggcgtgctcaccccc
 P  K  S  H  S  R  Q  A  S  T  D  A  G  S  G  G  V  L  T  P
caccacgtacgcgcacactcgtctccagcttccctgcagctgggcgccgtttcgggtggc
 H  H  V  R  A  H  S  S  P  A  S  L  Q  L  G  A  V  S  G  G
tccctgtctgggatggcctcagcaggcgcctccccccaacatctccggcagtcgtcctac
 S  L  S  G  M  A  S  A  G  A  S  P  Q  H  L  R  Q  S  S  Y
gagattcccgacgacgtgcccctgccgccaggctgggagatggccaagaccagctccggc
 E  I  P  D  D  V  P  L  P  P  G  W  E  M  A  K  T  S  S  G
cagagatactttctcaaccacattgatcaaaccaccacttggcaggatccacgcaaagcc
 Q  R  Y  F  L  N  H  I  D  Q  T  T  T  W  Q  D  P  R  K  A
ctgctccagttgaaccaggctacccctcccagtaccgttccagttcagcagcagaacctt
 L  L  Q  L  N  Q  A  T  P  P  S  T  V  P  V  Q  Q  Q  N  L
ttaagcccagctagtggacctctacctgaaggctgggaacaagccataacaccagaggga
 L  S  P  A  S  G  P  L  P  E  G  W  E  Q  A  I  T  P  E  G
gaaatttactacattaaccacaagaacaagactacttcctggctagatcctcgacttgag
 E  I  Y  Y  I  N  H  K  N  K  T  T  S  W  L  D  P  R  L  E
actcgctatgcc
 T  R  Y  A
```

II-e Medaka TAZ full-length sequences

```
atggacgcgcaccggagcggacggcaggtcgtgcacgtccgcggggactcgcagaaggag
 M  D  A  H  R  S  G  R  Q  V  V  H  V  R  G  D  S  Q  K  E
ctggaggccctgttcagtagggtgatgaaccccagcgaggcgtccaggcagccgtcgtcg
 L  E  A  L  F  S  R  V  M  N  P  S  E  A  S  R  Q  P  S  S
gttccgatgaggatgaggaagctgcccgactccttcttcaagcccccggagccccgcgga
 V  P  M  R  M  R  K  L  P  D  S  F  F  K  P  P  E  P  R  G
cactctagacaagccagttcggatggaggcgtgtgtggttccctcactccccatcacagc
 H  S  R  Q  A  S  S  D  G  G  V  C  G  S  L  T  P  H  H  S
cggacggtttccgcccccgctgcgctccccgtcaactccctctccacccaagcagcagat
 R  T  V  S  A  P  A  A  L  P  V  N  S  L  S  T  Q  A  A  D
gtcgcggcgacatcgatcatacctgacgatgtgccactccccccctggttgggaaatggcc
 V  A  A  T  S  I  I  P  D  D  V  P  L  P  P  G  W  E  M  A
aaaactcccgctggccagcgatacttcctgaatcacctggataagacaaccacgtggctc
 K  T  P  A  G  Q  R  Y  F  L  N  H  L  D  K  T  T  T  W  L
gaccctcgtctctcgcagctccagtcggctcagcatcccatcgccaacgccccggtccat
 D  P  R  L  S  Q  L  Q  S  A  Q  H  P  I  A  N  A  P  V  H
tcccactccttcagcaaccctgcgcccaccacacaagcacaaagcgtcaacccggagtca
 S  H  S  F  S  N  P  A  P  T  T  Q  A  Q  S  V  N  P  E  S
ggtccactgcctgaaggctgggagcaggctgtgactccagagggagagatgtactacatc
 G  P  L  P  E  G  W  E  Q  A  V  T  P  E  G  E  M  Y  Y  I
gatcacataaataagaccaccacatgggtcgacccgcgtctagctcagaaaatgagccct
 D  H  I  N  K  T  T  T  W  V  D  P  R  L  A  Q  K  M  S  P
ggtgtcctcagcttggcaatgcaaaggcaggaaaagctcagatgcaagcaaggcgcgccc
 G  V  L  S  L  A  M  Q  R  Q  E  K  L  R  C  K  Q  G  A  P
ccacaagccacacaacaggagggagcaggaaggaaccaaatgtctgctggtttggaccac
 P  Q  A  T  Q  Q  E  G  A  G  R  N  Q  M  S  A  G  L  D  H
gacaggagcggtcaggtacttgtcccatcagtggatgccaggatccgagccccaaaccat
 D  R  S  G  Q  V  L  V  P  S  V  D  A  R  I  R  A  P  N  H
gaaactgcaatgaacggagctcactctcgcaatgagagcacggacagcggtctgagcgtc
 E  T  A  M  N  G  A  H  S  R  N  E  S  T  D  S  G  L  S  V
agcagcctgccgcgcacgaacgaccacatgctgagccctgtggatcacatggatactggt
 S  S  L  P  R  T  N  D  H  M  L  S  P  V  D  H  M  D  T  G
gaccccagtgaaccccccatcgatggccctgcaggagccgatgctccccatgagtgaaggc
 D  P  S  E  P  P  S  M  A  L  Q  E  P  M  L  P  M  S  E  G
gaggagctgatgccctgcatccctgaaggtctgagctcagacctcctgatggatatggag
 E  E  L  M  P  C  I  P  E  G  L  S  S  D  L  L  M  D  M  E
actgttctctccggctcacacatggacagggacaatctgctcacctggctataggc
 T  V  L  S  G  S  H  M  D  R  D  N  L  L  T  W  L  -
```

II-f Drosophila *Sav (WW1-WW2-SARAH: pSV281) sequences*

```
atgggttcttctcaccatcaccatcaccatgaaaacctgtacttccagggatccgaggag
 M  G  S  S  H  H  H  H  H  H  E  N  L  Y  F  Q  G  S  E  E
ctgcctctgccgcccggctgggccactcagtacacgctacacggtcggaaatactatatt
 L  P  L  P  P  G  W  A  T  Q  Y  T  L  H  G  R  K  Y  Y  I
gatcacaatgcgcataccacgcactggaatcatccgttggagcgcgaaggtctgccggtg
 D  H  N  A  H  T  T  H  W  N  H  P  L  E  R  E  G  L  P  V
ggctggcggcgggtggtgtccaagatgcatggcacctactatgagaaccagtataccggg
 G  W  R  R  V  V  S  K  M  H  G  T  Y  Y  E  N  Q  Y  T  G
cagagccaacgtcagcatccatgcttgacctcctactatgtctacacgacgtctgcggag
 Q  S  Q  R  Q  H  P  C  L  T  S  Y  Y  V  Y  T  T  S  A  E
ccaccgaaagcgattcgaccagaggcgtcgctctatgccccacccacgcacactcacaat
 P  P  K  A  I  R  P  E  A  S  L  Y  A  P  P  T  H  T  H  N
gcactggtgccggccaatccctatctgctcgaggagatccccaagtggttggccgtctac
 A  L  V  P  A  N  P  Y  L  L  E  E  I  P  K  W  L  A  V  Y
tcggaggcggactcgtccaaggaccacctgctgcagttcaacatgtttagcctgccggag
 S  E  A  D  S  S  K  D  H  L  L  Q  F  N  M  F  S  L  P  E
ctggagggcttcgacagcatgctggtgcggctcttcaagcaggaactgggcaccatcgtg
 L  E  G  F  D  S  M  L  V  R  L  F  K  Q  E  L  G  T  I  V
ggcttctacgagcgctaccgtcgcgctttgatactcgagaagaatcgacgcgccggccag
 G  F  Y  E  R  Y  R  R  A  L  I  L  E  K  N  R  R  A  G  Q
aaccagaaccaaaaccagtga
 N  Q  N  Q  N  Q  -
```

II-g Human Sav (WW1-WW2-SARAH: pSV281) sequences

```
atgggttcttctcaccatcaccatcaccatgaaaacctgtacttccagggatccgaagat
 M  G  S  S  H  H  H  H  H  H  E  N  L  Y  F  Q  G  S  E  D
ttacccttcctcctggctggtctgtggactggacaatgagagggagaaaatattatata
 L  P  L  P  P  G  W  S  V  D  W  T  M  R  G  R  K  Y  Y  I
gatcataacacaaatacaactcactggagccatcctcttgagcgagaaggacttcctcct
 D  H  N  T  N  T  T  H  W  S  H  P  L  E  R  E  G  L  P  P
ggatgggaacgagttgagtcatccgaatttggaacctattatgtagatcacacaaataag
 G  W  E  R  V  E  S  S  E  F  G  T  Y  Y  V  D  H  T  N  K
aaggcccaatacaggcatccctgtgctcctagtgtacctcggtatgatcaaccacctcct
 K  A  Q  Y  R  H  P  C  A  P  S  V  P  R  Y  D  Q  P  P  P
gtcacataccagccacagcaaactgaaagaaatcagtcccttctggtacctgcaaatcca
 V  T  Y  Q  P  Q  Q  T  E  R  N  Q  S  L  L  V  P  A  N  P
tatcatactgcagaaattcctgactggcttcaggtttacgcacgagcccctgtgaaatat
 Y  H  T  A  E  I  P  D  W  L  Q  V  Y  A  R  A  P  V  K  Y
gaccacattctgaagtgggaactcttccagctggctgacctggatacataccagggaatg
 D  H  I  L  K  W  E  L  F  Q  L  A  D  L  D  T  Y  Q  G  M
ctaaagttgctcttcatgaaagaattggagcagattgttaaaatgtatgaagcatacaga
 L  K  L  L  F  M  K  E  L  E  Q  I  V  K  M  Y  E  A  Y  R
caagcccttcttacagagttggaaaaccgaaagcagagacagcagtggtatgcccaacaa
 Q  A  L  L  T  E  L  E  N  R  K  Q  R  Q  Q  W  Y  A  Q  Q
catggatga
 H  G  -
```

II-h Human Kibra (WW1-WW2 long: pBG101) sequences

```
caccatgaaaacctgtacttccagggatccccgaattccgagctgcccctgccggag
 H  H  E  N  L  Y  F  Q  G  S  P  N  S  E  L  P  L  P  E
ggctgggaggaggcgcgcgacttcgacggcaaggtctactacatagaccacacgaaccgc
 G  W  E  E  A  R  D  F  D  G  K  V  Y  Y  I  D  H  T  N  R
accaccagctggatcgacccgcgggacaggtacaccaaaccgctcacctttgctgactgc
 T  T  S  W  I  D  P  R  D  R  Y  T  K  P  L  T  F  A  D  C
attagtgatgagttgccgctaggatgggaagaggcatatgacccacaggttggagattac
 I  S  D  E  L  P  L  G  W  E  E  A  Y  D  P  Q  V  G  D  Y
ttcatagaccacaacaccaaaaccactcagattgaggatcctcgagtacaatggcggcgg
 F  I  D  H  N  T  K  T  T  Q  I  E  D  P  R  V  Q  W  R  R
gagcaggaacatatgctgaagtaa
 E  Q  E  H  M  L  K  -
```

Appendix III: Buffers and solutions

III-a 100 ml M9 Trace elements

$CaCl_2.2H_2O$	550 mg
$MnSO_4.H_2O$	140 mg
$CuSO_4.5H_2O$	40 mg
$ZnSO_4.7H_2O$	220 mg
$CoSO_4.7H_2O$	45 mg
$Na_2MoO_4.2H_2O$	26 mg
H_3BO_4	40 mg
KI	26 mg

-Add 70 ml distilled H_2O and adjust to pH8

-Add 500 mg EDTA and adjust to pH8

-Add 375 mg $FeSO_4.7H_2O$

Make up to 100 ml with distilled H_2O and autoclave

III-b 15% acrylamide gel for SDS-PAGE (8 gels)

Resolving gel

Tris base 1.5 M pH 8.8	12.5 ml
H_2O	18.5 ml
40% 29:1 acrylamide/bis-acrylamide	18.7 ml
10% APS	250 µl
TEMED	22.5 µl

-Pipette saturated butanol onto the liquid resolving gel and remove using filter paper (Whatman) before pouring the stacking gel.

Stacking gel

Tris 0.5 M pH6.8	3.125 ml
H_2O	8 ml
40% 29:1 acrylamide/bis-acrylamide	1.25 ml
10% APS	62.5 µl
TEMED	12.5 µl

III-c 10 ml 2x SDS-PAGE sample buffer

H_2O	2.25 ml
0.5 M Tris pH6.8	1.25 ml
10% (w/v) SDS	3 ml
50% (v/v) glycerol	2 ml
2-mercaptoethanol	0.5 ml
0.15% bromophenol blue	1 ml

III-d 1 L 10x Glycine running buffer for SDS-PAGE

Tris base	30.3 g
Glycine	143.75 g
SDS	10 g

- Make up to 1 L with distilled H_2O. The pH is not adjusted but should be around pH 8.3.

III-e Coomassie Blue stain for acrylamide gels

40% Methanol

0.1% (w/v) Brilliant blue R250

50% H_2O 10% Acetic acid

- Dissolve brilliant blue R250 in methanol before adding the H_2O and acetic acid.

III-f Destain for acrylamide gels

25% Methanol

65% H_2O

10% Acetic acid

III-g Drying solution for acrylamide gels

30% Methanol

10% Glycerol

50% H_2O

10% Acetic acid

III-h 1 L 50x TAE buffer for DNA electrophoresis

Tris base	242 g
Acetic acid	57.1 ml
0.5 M EDTA	100 ml

- EDTA solution must be adjusted to pH 8 to enhance the solubility of the EDTA powder.
- Make up to 1 L with distilled H_2O

III-i Buffers for His-affinity purification

His buffer A (wash buffer) – pH 8

20 mM Tris base

300 mM NaCl

20 mM Imidazole

His buffer B (elution buffer) – pH 8

20 mM Tris base

300 mM NaCl

500 mM Imidazole